Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

Control charts for autocorrelated processes under parameter estimation

A dissertation presented by

Jorge Arturo Garza Venegas

Submitted to the
School of Engineering and Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Engineering Science

Major in Industrial Engineering

Monterrey, Nuevo León, December 4th, 2018

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

The committee members, hereby, certify that have read the dissertation presented by Jorge Arturo Garza Venegas and that it is fully adequate in scope and quality as a partial requirement for the degree of Doctor of Philosophy in Engineering Science, with a major in Industrial Engineering.

Dr. Víctor Gustavo Tercero Gómez
Tecnológico de Monterrey
School of Engineering and Sciences
Principal Advisor

Dr. Philippe Castagliola
Université de Nantes
LS2N UMR CNRS 6004
Co-Advisor

Dr. Francisco Román Ángel Bello Acosta
Tecnológico de Monterrey
Committee Member

Dr. Alvaro Eduardo Cordero Franco
Universidad Autónoma de Nuevo León
Committee Member

Dr. Neale Ricardo Smith Cornejo
Tecnológico de Monterrey
Committee Member

Dr. Rubén Morales Menéndez
Associate Dean of Graduate Studies
School of Engineering and Sciences

Monterrey, Nuevo León, December 4th, 2018

Declaration of Authorship

I, Jorge Arturo Garza Venegas, declare that this dissertation titled, "Control charts for autocorrelated processes under parameter estimation" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this dissertation has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work.
- I have acknowledged all main sources of help.
- Where the dissertation is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Jorge Arturo Garza Venegas
Monterrey, Nuevo León, December 4th, 2018

Dedication

To my parents: Minerva Venegas Chavarría and Oscar Garza Mora, for their unconditional love and support.

Acknowledgments

First of all, I would like to express my deepest gratitude to my family, especially to my parents Minerva and Oscar for their unconditional love, the most amazing feeling I have experienced; to my brothers “Heri” and “Rola” for all the years we shared growing up together. To Brenda for giving me the bliss of being the uncle of, until now, four girls. I love all of you.

I want to thank my advisor, Dr. Víctor Gustavo Tercero Gómez, for his support over these years. I appreciate all the time you spent working with me and giving me advice on being a better person, professor, and researcher. I really admire the way you work and I am so proud and happy to have done my Ph. D. under your supervision. You are one of the persons I admire the most. I think you are an exceptional leader.

Thanks to Dr. Philippe Castagliola for accepting me as a visiting student at *Institut Universitaire de Technologie de Nantes* and accepting to be my co-advisor. It was an amazing experience working with you in France, both academic and personal. I really admire and like the way you work with the students. You are so concise and focus on the really important things, which helped me to not get lost on all details but just on the important ones. I am so pleased and proud to have you as my co-advisor. *Merci beaucoup !!*

Thanks to the committee members Dr. Eduardo Cordero, Dr. Francisco Ángel-Bello and Dr. Neale Smith for all the corrections made on the document and research that increased the value of the work I carried out during these years.

I am grateful to Dr. Giovanni Celano for his constructive comments and suggestions about Chapters 3 and 4. These have improved the research for publication in a journal. Thanks to Dr. Linda Lee Ho for her suggestions when extending results from Chapter 3 to a journal paper.

Thanks to all my friends all around the world. It is impossible for me to mention all of them here. They made these years more enjoyable and less frustrating.

Last but not least, thanks to the Tecnológico de Monterrey for the support on tuition and to the CONACyT for the support for living. Without that help, I would have not been able to complete my doctoral studies successfully neither been able to make my research stay in France.

Control charts for autocorrelated processes under parameter estimation.

By

Jorge Arturo Garza Venegas

Abstract

Statistical Processes Monitoring is a collection of statistical-based methodologies and methods for monitoring the quality of manufactured products or services. Within these tools, control charts are powerful ones to assist practitioners on the detection of departures from in-control situations as long as the assumptions made on their design are fulfilled; otherwise, their power might decrease. For instance, control charts performance has been shown to be negatively affected when using estimated parameters (in which case the Average Run Length, ARL, becomes a random variable) or when dealing with autocorrelated data. Given that, this research is focused on the effect of parameter estimation on the performance of the $\bar{X}$ and the modified S^2 control charts for monitoring the mean and the variance, respectively, of autocorrelated processes under parameter estimation. The average of the ARL and its standard deviation are considered as performance measures as they take into account the sampling variability of the ARL. Furthermore, a bootstrapping methodology is applied to adjust control limits in order to have a guaranteed conditional in-control performance with a certain probability and the effect on the out-of-control ARL is also studied.

List of Figures

List of Tables

Contents

Chapter 1. Introduction

1.1 History and background

One common task done when considering manufacturing processes is their monitoring, which allows to determine whether or not the products obtained meet some constraints related to their quality. It is usually desired to obtain products as similar to a quality standard as possible, however, Shewhart (1931) stated that there is always a variation when doing the same task several times, and that this variation could be considered as a natural (or inherent to the process) variation, due to common causes or it could be due to some specific causes, called assignable causes. Furthermore, he provided the basis of what is known as Statistical Process Control (SPC) (recently, Statistical Process Monitoring (SPM)) founded on several postulates, within which the third one states: “Assignable causes of variation may be found and eliminated” (Shewhart, 1931). For such purpose, SPM provides statistical methods or tools to help practitioners in the searching of assignable causes of variation, being the control charts one of the most widely used ones due to their simplicity and easiness to be implemented as well as their power to detect departures from in-control situations. Control charts were originally developed by Shewhart and since their creation, other control charts have been proposed as the Cumulative Sum (CUSUM) chart of Page (1954) and the Exponentially Weighted Moving Average (EWMA) chart introduced by Roberts (1959), which are preferred over Shewhart control charts when the main concern is the detection of small shifts in the mean.

Despite their usefulness, control charts are designed by considering certain assumptions about the process under monitoring and therefore their performance is tightly related to the compliance of them. Common assumptions include a specific parametric model or distribution for observations; uncorrelated/independent data; prior knowledge of the true values of the in-control parameters, among others. Moreover, the fact that these assumptions are seldom met in practice make the charts no longer suitable to be used for searching assignable causes of variation as their behaviour is unpredictable. For instance, Shewhart charts are designed for independent observations when in-control process parameters are known beforehand, leading to a well-known chart performance which usually is quite different than the one observed under correlated observations and/or estimated parameters, as summarized in Jensen et al. (2006) and Psarakis et al. (2014) literature reviews which include several types of control charts.

Control charts performance is related to the Run Length (RL), which is defined as the number of observations obtained before a signal is triggered by the chart, and which is a random variable. Charts performance has been usually measured in terms of the Average Run Length, ARL, the expected number of observed points within control chart limits before the chart triggers a signal. It is called in-control ARL, denoted by ARL_0, or out-of-control ARL, denoted by ARL_1 according to whether the process is under statistical control or not, respectively. Nevertheless, these performance measures might not be suitable to characterize the chart performance (for instance when the RL's distribution is skewed or heavy-tailed) and Jones-Farmer et al. (2004) suggested to report not only the ARL but also other RL distribution characteristics, including its standard deviation $SDRL$ or even some quantiles of its distribution. For Shewhart control charts where the statistic used for running the chart turns out to be independent, the RL has a geometric distribution and thus, the ARL could be obtained easily. On the other hand, Brook and Evans (1972) developed a Markov Chain approach to obtain approximations of the RL distribution for CUSUM control charts that was adapted by Lucas and Saccucci (1990) for EWMA control charts.

As some assumptions made on charts' design are rarely met in practice, several approaches to assess those situations have been proposed allowing the use of control charts for process monitoring. For instance, one of the main issues that have taken the attention of several researchers is the unknown parameters case. One approach to address this situation is to collect an assumed in-control sample (called Phase I sample, Chakraborti et al., 2009), to use it to estimate the process' parameters and use those estimations as process parameters true values to run the chart. Nevertheless, charts might no longer have the expected performance due to the estimation variability (Ghosh et al., 1981) and therefore, their use might not be longer suitable. In order to diminish that variation, it is suggested to collect a certain number of Phase I samples in order to have a more precise estimation, since the closer the estimation to the true value, the closer the chart performance to the known parameter performance. When the latter approach is not feasible (for instance, when dealing with start-up processes or short runs with few or not all historical data), the self-starting methodology proposed by Hawkins (1987) or the Q-statistics introduced by Quesenberry (1991a) are recommended since these approaches avoid the necessity of large initial in-control samples as parameters estimations are done from the very beginning of the process and are updated as more data is available.

Another assumption often made is the independence of the observations which not always could be ensured, as practical situations where autocorrelation arises are found in health surveillance, crop monitoring, chemical processes, etc.; where observations are collected from the same object or from continuous streams of data or within small periods of time. In fact, control charts are sensitive to the independence assumption as it has been shown by Johnson and Bagshaw (1974), Bagshaw and Johnson (1975) and Alwan (1992), among others. Nonetheless, there are several approaches to deal with this situation, including removing the autocorrelation between observations by using different sampling techniques (e.g. Franco et al., 2014), modification of control the control limits (e.g. Costa and Castagliola, 2011) and the time series approach where the residuals of a fitted time series model are monitored instead of the observations (e.g. Alwan and Roberts, 1988). The latter approach leads to the design of control charts called *residuals* control charts and it has been widely used since it allows to handle several autocorrelation structures. However, main drawbacks of implementing residuals charts are the necessity of prior knowledge of the model (to ensure residuals are independent) and the fact that a change in the observations might not be captured by residuals, affecting their performance (Longnecker and Ryan, 1991). As an alternative for residuals charts, several control charts for stationary processes have been developed (e.g. Alshraideh and Khatatbeh, 2014; Zhang and Pintar, 2015; Franco et al., 2015; Osei-Aning et al., 2017a).

Regardless the fact that these approaches allow the use of control charts, the non-compliance of the assumptions made on their design make the RL distribution behaves quite different as expected. For instance, under the known parameters case the ARL is a constant in opposition to the estimated parameters case, where the ARL becomes a random variable, and therefore, making the previous performance measures not suitable for evaluating charts performance. The rationale is that the ARL will be conditioned on the estimates obtained during Phase I and used to design the control chart: two different random samples taken from the very same process will lead to different control limits and therefore different ARL's. The practical implication is that two practitioners could (and probably) observe quite different performance when implementing the same control chart to data gathered from the same process distribution, leading to an uncertainty about the effectiveness in implementing the chart for the search of assignable causes of variation (the false-alarm rate might be inflated or the ability to detect true changes might be low). The ARLs sampling variation was recognized by Jones and Steiner (2012) and some authors called it as “practitioner-to-practitioner” variability (Zhang et al., 2014, Saleh et al., 2015a, Saleh et al., 2015b) and they proposed to use the ARL's distribution to characterize the charts performance under parameter estimation, being the $AARL$ and $SDARL$ measures commonly reported when evaluating control chart performance under estimated parameters.

The amount of data gathered from Phase I samples in practice turns out to be insufficient to achieve the known parameter case performance when the “practitioner-to-practitioner” variability is considered. As a rule of thumb Zhang et al. (2014) suggested to have an $SDARL_0$ within the 10% of the nominal ARL_0 value. Therefore, Shewhart, CUSUM and EWMA control charts have been reevaluated considering their in-control conditional performance and it was concluded the necessity of quite large (and practically unrealistic) amounts of Phase I data to achieve Zhang et al.'s suggestion. However, Gandy and Kvaløy (2013) developed a methodology in order to obtain a guaranteed conditional in-control performance (with certain probability) when designing control charts, consisting on the adjustment of the control limits via bootstrapping methods. The main advantage of this methodology is that it is applicable to Shewhart, CUSUM and EWMA charts and that the control limits adjustment has been proved to have a moderate impact on the ARL_1.

To sum up, two main issues arising when implementing a control chart are the parameters estimation and the correlated observations since charts performance is affected negatively and lead to an uncertainty about their effectiveness to detect out-of-control situations. Given that the independence assumption is not always suitable and that the in-control parameters are often unknown, it is desirable to have a process monitoring tool capable to deal with both issues effectively in terms of charts performance, which have to take into account the "practitioner-to-practitioner" variation due to the Phase I estimations. Therefore, this research is focused on the evaluation of the performance of control charts for autocorrelated processes under parameters estimation, particularly the well-known $\bar{X}$ and S^2 control charts for monitoring the mean and the variance, respectively, of autoregressive processes or order 1, AR(1), in order to see if: (1) these well-known tools used on the independent case are good enough to monitor correlated processes or (2) they could be easily adapted to deal with this kind of processes or (3) there is a real necessity to develop new tools. The next subsection is devoted to the motivation for this research with the advantages and drawbacks of previous proposed approaches whereas Chapter 2 contains more details about these approaches.

1.2 Motivation

Even though control charts are powerful tools for process monitoring, they should not be implemented carelessly in practice as pointed out previously. For instance, for the Shewhart control charts under known parameters the RL has a geometric distribution with parameter α, where α is the probability of a false alarm. As control limits are set to be $\pm 3\sigma$ from the process mean μ, therefore the interval $[\mu - 3\sigma, \mu + 3\sigma]$ covers approximately the 99.73% of the population and then $\alpha = 0.0027$. Therefore, $ARL_0 = \frac{1}{\alpha} \approx 370.4$.

The previous statement means that when applying a Shewhart control chart to an in-control normally distributed process it is expected that the chart will trigger a signal every 370 observations. Now, applying the previous Shewhart control chart design to an AR(1) process with unknown parameters could lead to a different situations illustrated in Figures 1.2.1 to 1.2.4, where control limits were calculated considering the estimated mean and variance from an initial in-control sample of 100 observations from an AR(1) process and after that this chart is used for process monitoring. In Figure 1.2.1, the Shewhart control chart was applied to a one realization of the AR(1) process with autocorrelation coefficient $\phi = 0.9$ leading to 57 signals in a series of 370 in-control observations whereas in Figure 1.2.2, a sustained shift in the mean of 2 standard deviations at the 185-th observation is missed by the chart (even when it is noticeable by the naked eye). Considering negative values of ϕ, Figure 1.2.3 is the Shewhart control chart applied to a one realization of an in-control AR(1) process with $\phi = -0.9$ where there is no signal in 370 observations whereas Figure 1.2.4 is the chart applied to an AR(1) process with a shift in the mean of 1.5 standard deviations at the 185-th observation which is missed by the chart.

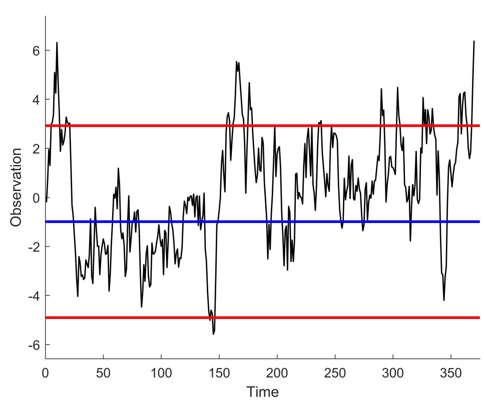


Figure 1.2.1: Example of a Shewhart control chart applied to an in-control AR(1) process with $\phi = 0.9$. The chart signals 57 times of 340.

From Figures 1.2.1 and 1.2.3 one might think that the solution is to widen control limits in order to take into account the variance of the process to avoid false alarms. On the other hand, Figures 1.2.2 and 1.2.4

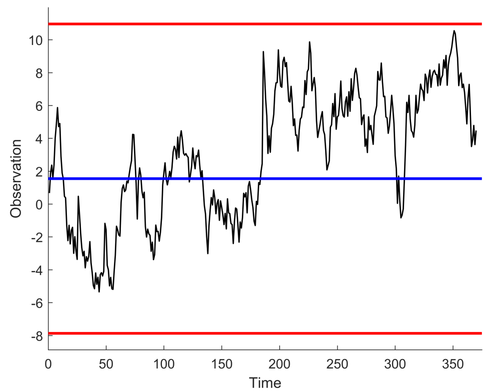


Figure 1.2.2: Example of a Shewhart control chart applied to an AR(1) process with $\phi = 0.9$ with a shift in the mean of 2 standard deviations at the 185-th observation.

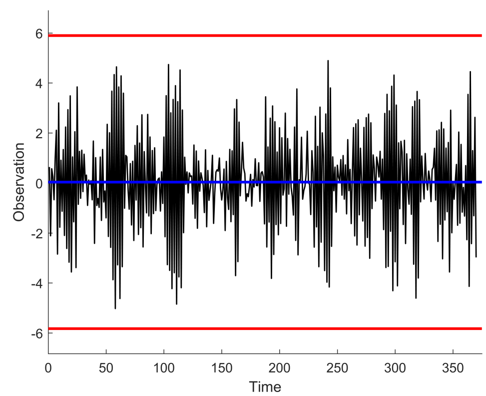


Figure 1.2.3: Example of a Shewhart control chart applied to an in-control AR(1) process with $\phi = 0.9$. No one signal in 370 runs.

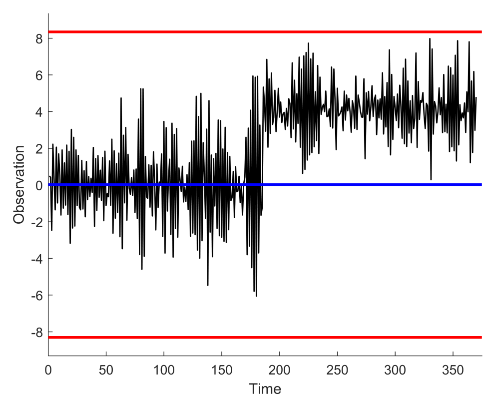


Figure 1.2.4: Example of a Shewhart control chart applied to an AR(1) process with $\phi = -0.9$ with a shift in the mean of 1.5 standard deviations at the 185-th observation.

suggest that this widening should not be made carelessly, since shifts in the process mean of 2 and 1.5 standard deviations, respectively, are missed by the control chart. Even though the autocorrelation coefficients used here are for high correlation levels, this kind of processes are found in literature and might be of more interest since low to moderate levels of autocorrelation may be addressed with traditional techniques.

The issue of parameters estimation has been addressed for the independent and identically distributed (i.i.d.) case considering different approaches, each one with their own advantages and disadvantages. For example, using parameters estimates as the true vales allows to implement a traditional control chart as in the known parameters case. Nevertheless, doing this always involves an estimation error which makes the chart underperforming or having an unexpected behaviour. After that, the efforts were focused on the determination of the estimators to be used and on the initial in-control sample size required to diminish the estimation error in order to the chart to behave as in the known case. As this approach is not feasible when sampling costs are highly expensive and/or when dealing with start-up processes the self-starting methodology and the Q-charts were developed. The main idea of these approaches is to transform the collected data to independent normal standard random variables and to use control charts on the transformed data. However, the magnitude of a change on the original variables might not be preserved on the transformed data and also biased charts (charts which the property that the probability to correctly signal an out-of-control situation is less than the probability to give a false alarm) could be obtained. In such cases, a slight modification on the chart leads to a better chart performance, as suggested by He et al. (2008) for the Shewhart Q-charts. These approaches are summarized on Table 1.2.1.

Table 1.2.1: Summary of solution approaches for the i.i.d. case under parameter estimation.

Solution approach	Advantages	Drawbacks
Use estimates.	Allows use of traditional charts.	Performance is negatively affected due to the estimation error.
Use of a (sufficiently) large initial in-control sample.	Diminishes the estimation error, and thus, the effect on chart performance due to estimation.	Not suitable for processes without historical data or expensive sampling costs.
Self-starting methodology	Avoid the necessity of Phase I samples to run the chart.	Provides biased charts.
Modified self-starting methodology.	Improve Self-starting control charts ($SSCC$) performance.	Formulation and/or implementation might be difficult to understand for practitioners/managers.

When dealing with autocorrelared streams of data, the first advice is to remove the autocorrelation structure of the process, which could be done by sampling less frequently or considering the variable sample interval (VSI) technique. However, there are several processes or situations with an inherent or natural autocorrelation structure where another approaches could be used such as taking into account the information given by the correlation structure and use it to modify charts design, e.g. by adjusting the control chart limits taking into account the process variation. One approach widely used is the time series approach, were data is fitted to a time series model where errors are assumed to be independent and then residuals of the fitted model are monitored with a control chart for independen observations. These charts are called *residuals* control chart. This approach allows to handle a wide variety of correlations since errors of the model are assumed to be independent; but, as in the i.i.d. case model, parameters are rarely known and have to be estimated and residuals (used as estimators of the errors) are not independent. In addition to this, a change in the observations might not be captured by residuals, as it is the case with the transformations involved on the self-starting control charts. Finally, several control charts have been proposed for stationary processes but there is a lack of studies about their performance, in terms of $SDARL$. Approaches used to assess the issue of correlated observations are summarized in Table 1.2.2.

Table 1.2.2: Summary of solution approaches for the autocorrelated data case.

Solution approach	Advantages	Drawbacks
Remove the autocorrelation.	Allows the use of traditional charts.	Not all available information is used. Might not apply to several processes.
Modification of control limits.	Charts take into account the autocorrelation.	Performance might be negatively affected due to parameters estimation.
Time series	Allows to handle a wide variety of correlations.	Performance relies on the complete knowledge of the process model.

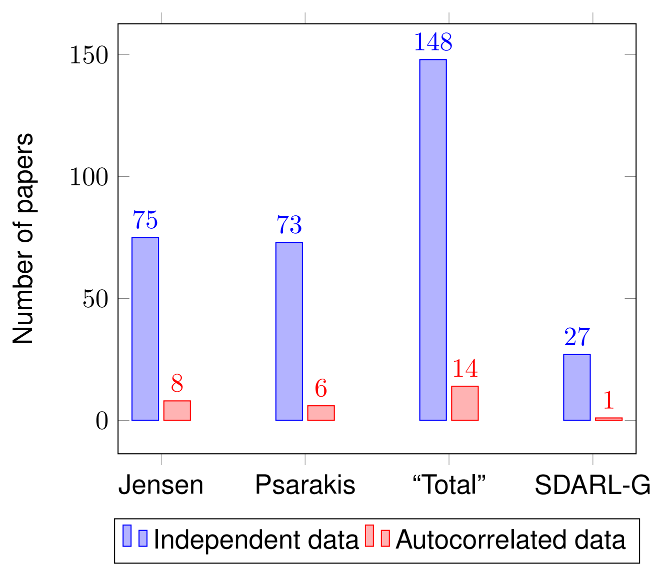


Figure 1.2.5: Number of publications about control charts for autocorrelated data under estimated parameters.

In addition to the problem of estimation, charts performance has been usually measured in terms of ARL and RL moments instead of using $AARL$ and $SDARL$ since the ARL is a random variable when parameters are estimated. In fact, recently several authors have been reevaluated control charts performance under estimated parameters using the $AARL$ and $SDARL$ as performance measures as they render better what happens in practice; for instance, Saleh et al. (2015b) and Goedhart et al. (2017b) reevaluated Shewhart control charts; Keefe et al. (2015), self-starting control charts; Saleh et al. (2015a) and Aly et al. (2016), EWMA control charts; and Saleh et al. (2016), CUSUM control charts, to mention a few (more information can be found in Chapter 2). In addition to these studies, head-to-head comparisons between control charts have been done: Hawkins and Wu (2014) considered the comparison under known parameters case whereas Zwetsloot and Woodall (2017) considered the unknown parameters case. General conclusions include the necessity of larger (in some cases unrealistic) amounts of Phase I data to achieve the known parameter case performance and that the EWMA and CUSUM control charts performance are not similar under parameters estimation, as was believed before. However, all these studies are devoted to the independent case and there has not been found works for their autocorrelated counterpart.

As shown above even the time series approach for handling autocorrelated data have to deal with parameter estimation in addition to the complexity of the model selection process and the fact that changes might not be captured by residuals. Moreover, there is a gap on the evaluation of control charts for autocorrelated processes under parameters estimation and/or with a guaranteed conditional in-control performance as can be seen in Figure 1.2.5. In that figure, the number of papers found in Jensen et al. (2006) (who considered papers from 1939 to 2005) and Psarakis et al. (2014) (who updated the studies up to 2013) literature reviews are shown in blue for the i.i.d. case and in red for the case of autocorrelated data. The column “Total” stands for the sum of the number of papers within this two literature reviews and the last column for the number of papers devoted to the conditional in-control performance and/or the guaranteed conditional performance (labeled as “SDARL-G”).

There is also a need of having tools easy to be implemented like the Shewhart control charts. Given these points, this research is focused on the study of the effect of parameters estimation on the modified $\bar{X}$ and S^2 control charts for autocorrelated processes using the $AARL$ and $SDARL$ as performance measures. The rationale of doing this is to have a better understanding of charts performance under parameter estimation and to have evidence about whether or not these tools are good enough to be implemented in practice by considering the new chart parameters that take into account the practitioner to practitioner variability or whether there is a necessity to improve them or even to develop new ones.

1.3 Problem Statement

Consider a sequence of observations $\{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\}$ distributed according to

- the following AR(1) model:

$$X_{i,j} - \mu_0 - \delta\sigma_0 = \phi_0 \left(X_{i,j-1} - \mu_0 - \delta\sigma_0\right) + \varepsilon_{ij} \tag{1.3.1}$$

  if the process mean is monitored, and

- the following model:

$$\frac{X_{i,j} - \mu_0}{\tau\sigma_0} = \phi_0 \left(\frac{X_{i,j-1} - \mu_0}{\tau\sigma_0}\right) + \varepsilon_{ij} \tag{1.3.2}$$

  if the process variance is monitored,

for $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, n$ where n is the subgroup size at the time i and where:

- $\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon\right)$ are uncorrelated.
- Process' parameters μ_0, σ_0, ϕ_0 are unknown.
- $\phi_0 \in (-1, 1)$ in order to have a stationary process.
- $\delta = \dfrac{|\mu_1 - \mu_0|}{\sigma_0}$ is the shift in the mean from μ_0 to μ_1 measured in standard deviations. Thus, $\delta = 0$ means no change in the mean.
- $\tau^2 = \dfrac{\sigma_1^2}{\sigma_0^2}$ measures variance shifts. Then, $\tau^2 = 1$ indicates no changes in variance.
- When $\delta = 0$ and $\tau^2 = 1$, the process is said to be under statistical control.
- $X_{i,0}$ is assumed to have the steady-state process distribution.
- Autocorrelation is assumed to be within not between samples: that is to say, $\forall j = 1, \ldots, n$, $X_{i,j}$ and $X_{i+1,j}$ are considered to be independent.

The AR(1) model in (1.3.1) is a widely known AR(1) model with mean $\mu_0 + \delta\sigma_0$ whereas the model shown in (1.3.2) is a modification of the traditional AR(1) model where the process variance is no longer related to the errors variance, as it will be seen in Chapter 5. The rationale of using model (1.3.2) is because assuming the steady-state distribution for all X_i's the changes in variance do not involve the errors variance, as it was usually done. However, a model similar to that in equation (1.3.1) could be used for variance monitoring, as they are equivalent.

This research addresses the problem of measuring the effect of the Phase I sample used to estimate parameters on the performance of mean and variance control charts for autocorrelated processes in terms of the $AARL_0$ and $SDARL_0$ and the corresponding effect on ARL_1 of using bootstrapping approaches to guarantee a minimum conditional in-control performance. The main goal is to evaluate the performance of control charts for monitoring AR(1) processes, particularly, the modified $\bar{X}$ control chart for monitoring the process mean, μ_0, and the modified S^2 control chart for monitoring the process variance, σ_0^2, under estimated parameters.

It is expected that the performance of control charts will be affected due to parameter estimation as in the i.i.d. case and as Figures 1.2.1 to 1.2.4 suggest. Moreover, as the widening of the control chart limits has not to be done carelessly, the bootstrapping methodology proposed by Gandy and Kvaløy (2013) where control limits are adjusted in order to guarantee a certain conditional in-control performance might be applied to these charts and the effect on ARL_1 should be studied.

1.4 Research questions

This research is a three-folded research consisting on the following studies:

1. The effect of the autocorrelation estimators on the $\bar{X}$ control chart performance
2. The conditional performance of the $\bar{X}$ control chart for AR(1) processes under estimated parameters
3. The conditional performance of the S^2 control chart for AR(1) processes under estimated variance,

where the first one is done in order to see the effect of autocorrelation estimators on the chart performance, for which the process mean and variance are considered as known and fixed and only the autoregressive parameter is unknown. After that, the next step is to assume all parameters as unknown. The third one is left for monitoring process variance, where the effect of autocorrelation as well as process variance estimation will be considered.

- Q_1.- What is the conditional performance of $\bar{X}$ control chart for monitoring the mean of AR(1) processes when using autocorrelation estimators?
- Q_2.- What is the conditional performance of $\bar{X}$ control chart for monitoring the mean of AR(1) processes under parameter estimation?
- Q_3.- What is the conditional performance of the modified S^2 control chart for monitoring the variance of AR(1) processes when the variance is estimated?
- Q_4.- What is the effect on the performance (in terms of $AARL$ and $SDARL$) of $\bar{X}$ and S^2 control charts when applying the Gandy and Kvaløy's bootstrap methodology to adjust control limits to have a guaranteed conditional in-control performance?

First three questions are related to the effect of parameter estimation on the ARL of the $\bar{X}$ and S^2 control charts, which has not been addressed yet while, the fourth one, is non-trivial since Gandy and Kvaløy methodology assumes handling independent observations coming from a probability distribution function whereas for AR(1) processes observations follows *a model* instead of a probability distribution function alongside the fact that observations are not independent.

1.5 Research hypotheses

Considering the previous research questions, it is suspected that:

- H_1.- The performance, in terms of the $AARL_0$ and $SDARL_0$, of the $\bar{X}$ control chart is different from the known parameter case and it is negatively affected by using autocorrelation estimators.
- H_2.- The performance, in terms of the $AARL_0$ and $SDARL_0$, of the $\bar{X}$ control chart is different from the known parameter case and it is negatively affected under parameter estimation.
- H_3.- The performance, in terms of the $AARL_0$ and $SDARL_0$, of the modified S^2 control chart is different from the known variance case and it is negatively affected under variance estimation.
- H_4.- When applying Gandy and Kvaløy's methodology to $\bar{X}$ and S^2 control charts for AR(1) processes the effect on the ARL_0 is as desired and on ARL_1 is not so severe.

The general hypothesis is that similar situations as in the i.i.d. case are going to arise: larger amounts of Phase I data have to be gathered in order to achieve the known parameter case performance, in some cases unrealistic amounts of data. Moreover, it is expected that this will be found specially in those cases with high degrees of autocorrelation. However, this could be addressed using the Gandy and Kvaløy methodology.

1.6 Research purpose and objective

Consider the process models stated in equations (1.3.1) and (1.3.2). The main purpose of the research is the evaluation of the effect of parameter estimation for autocorrelated processes following an AR(1) model in order to see if there is a necessity to develop a tool, a control chart, for managing processes in presence of autocorrelation and parameters estimation, and the implementation of bootstrapping approaches to guarantee a conditional in-control performance with certain probability. Even though there are several control charts that deal with the estimation problem, they were designed assuming an independent data structure and, similarly, control charts for autocorrelated processes usually assume prior knowledge of the in-control parameters, as it will be seen in the gap analysis of Chapter 2.

Therefore, the objectives of the research are:

- Evaluate the effect of estimating ϕ_0 on the conditional performance the $\bar{X}$ control chart for autocorrelated processes when only ϕ_0 is unknown. That is to say, study the relationship between the conditional ARL and $\hat{\phi}_0$.
- Evaluate the conditional performance of the $\bar{X}$ control chart for AR(1) processes when μ_0, σ_0 and ϕ_0 are unknown by means of $AARL$ and $SDARL$ and varying the number of Phase I data used to estimate process parameters.
- Evaluate the conditional performance of the modified S^2 control chart for AR(1) processes when σ_0^2 is estimated by means of $AARL$ and $SDARL$ and varying the number of Phase I data used to estimate the process variance.
- Show that it is possible to guarantee, with a certain probability, a conditional in control performance for the $\bar{X}$ and S^2 charts by adjusting their control limits carefully using a bootstrapping methodology and study the effect on the out-of-control performance.

1.7 Delimitations

Given the assumptions made in the models shown in equations (1.3.1) and (1.3.2), the following limitations arise:

- The random error is assumed to be distributed as $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$. Thus, other kind of noises are not considered on this study.

- Only AR(1) processes are considered in this study.
- The Average Run Length (ARL) distribution might not be obtained analytically, due to the lack of an exact or closed form of the chart statistic.
- Process model is assumed to be known *a priori*. Therefore, model selection procedures are not considered here.

Despite the limitations mentioned above, it is worthy to say that: the random errors are usually assumed to be distributed as $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$; that AR(1) processes could be used to represent a wide variety of real life processes and therefore, might set the basis to extend this studies to Moving Averages (MA), and Autoregressive and Moving Average (ARMA) processes; and that even when the ARL distribution might not be obtained analytically it is possible to obtain it by extensive Monte Carlo simulations.

1.8 Research outputs and outcomes

The major **output** of this research is the evaluation of the $\bar{X}$ and S^2 control charts performance for AR(1) processes under parameter estimation. There is not evidence of a previous study like this for autocorrelated processes, particularly for AR(1) processes. In that sense, better knowledge on the conditional performance of those control charts is expected to be obtained in this research. As other outputs of this research are (i) the calculation of the chart constants via bootstrapping methods in order to adjust the control charts limits, and (ii) a gap analysis making evident the lack of studies on SPM when handling autocorrelated data (and the exploration/modification of techniques proven to be useful for control charts for independent data).

On the other hand, the **outcome** is a procedure to assist in process management when dealing with autocorrelated processes without prior knowledge of the in-control process parameters, by considering the adjusted control limits to guarantee a conditional in-control performance with certain confidence, to be used in practice, that is to say, the implementation of these control charts under parameter estimation.

1.9 Dissertation organization

This dissertation follows a three paper style, so it is organized as follows: Chapter 2 gives the insights of the state of the art and summarizes the main contributions done in SPM, concluding with a gap analysis. Chapters 3, 4 and 5 are research papers that help to support the dissertation arguments: Chapter 3 is focused on the study of the effect of the autocorrelation estimators on the $\bar{X}$ control chart performance in terms of $AARL_0$ and $SDARL_0$, Chapter 4 is focused on the study of the performance of $\bar{X}$ control chart for monitoring the mean of AR(1) processes when all parameters are unknown, and Chapter 5 is left for the performance of the modified S^2 control chart for monitoring the variance of AR(1) processes when the process variance is estimated. Finally, conclusions and future work are left in Chapter 6.

Chapter 2. Background and literature review

2.1 Introduction

Shewhart (1931) set the scientific basis of quality control and provided statistical tools known as control charts (or Shewhart charts, named after him) for processes management. These tools were created with the main purpose of monitoring the process variation and determine whether or not it is under statistical control, which could boost the detection and elimination of special causes of variation in order to secure a state of control. The main advantages of Shewhart charts are their easiness to be understood and to be implemented whereas their main drawback is the fact that the assumptions made on their design are rarely met in practice, leading to unexpected, usually worst performance. Assumptions made on their design include the independence of observations, an underlying normal distribution and in-control process parameters known *a priori*, among others. Further, Shewhart charts are not the best option when concerns are the finding of small shifts in the process mean.

Given that there are some practical situations were the performance of the Shewhart control charts is not satisfactory or when the assumptions made on their design could not be validated or suitable to be used, several control charts or methodologies have been developed in order to overcome these situations. For instance, when dealing with small shifts in the mean, the Cumulative Sum (CUSUM) chart proposed by Page (1954) and the Exponentially Weighted Moving Average (EWMA) introduced by Roberts (1959) are better detecting those kind of shifts than Shewhart charts. Considering the parameter estimation case, it has been shown that the chart capability detection decreases due to the estimation error and that the performance measure common used, the Average Run Length (ARL) becomes a random variable instead of being a constant. Several approaches have been considered to assess this problem, being the increase of the Phase I sample size, the Q-transformation proposed by Quesenberry (1991a) and the self-starting methodology introduced by Hawkins (1987) the principal ones, all of them developed and used mainly for the i.i.d. case.

When dealing with autocorrelated streams of data, it has been proved that applying traditional control charts to these kind of data lead to chart underperformance. Solution approches to this issue include the modification of traditional control charts (Vasilopoulos and Stamboulis, 1978), the modification of the sampling technique (Reynolds et al., 1988) or the time series based control charts (as the *residuals* control charts introduced by Alwan and Roberts, 1988) or development of control charts for stationary processes. The time series approach is of particular interest since it allows to handle a wide variety of correlation structures such as the Autoregressive and Integrated Moving Average models (ARIMA) models.

Even though there are other issues when designing a control chart (such as the economic design or the economic-statistical design of a control chart) and as all Phase I methods are beyond the scope of this research (the reader is referred to the works of Chakraborti et al., 2009 and Jones-Farmer et al., 2014 for literature reviews about Phase I methods), this research is mainly focused on the parameter estimation problem alongside the autocorrelated data issue. Therefore, the next subsections are devoted to the presentation of the state of the art for the effect of parameter estimation on control charts performance, the self-starting methodology for the independent case and the SPM tools developed to deal with autocorrelated streams of data.

2.2 Effect of parameter estimation on control charts performance

When dealing with processes whose in-control parameters are unknown, the control charts applied to such processes tend to have a performance quite different to that expected under the known parameter case. Therefore, the majority of the works for this issue were devoted to have responses of the questions stated in Jensen et al. (2006):

> "Just how poorly (or maybe well) might a chart perform if designed with estimates in place of known parameters? What sample size is needed in Phase I to ensure adequate performance in

Phase II? How should the Phase II limits be adjusted to compensate for the size of the Phase I sample?".

or to develop alternatives to minimize the differences on the chart performance between known and unknown parameters cases.

The chart performance has been commonly studied by means of the Run Length (RL) which is a random variable that equals the number of plotted statistics before a signal is triggered from the control chart. The average of the RL (ARL) is commonly used to evaluate the chart performance and it is defined as the expected number of plotted statistics before a triggered signal from the chart. Considering the case of known parameters, as long as the chart statistics are independent and the control limits are fixed, which turns out to be the case of Shewhart charts, the RL has a geometric distribution with parameter α, the probability of a signal, and therefore, $ARL = \frac{1}{\alpha}$. Concerning the CUSUM control chart Brook and Evans (1972) developed a Markov Chain approach to calculate the ARL based on a discretization of the statistic values whereas Crowder (1987) considered an integral-equation for the same end but considering the EWMA control chart for which Lucas and Saccucci (1990) adapted the Markov Chain approach by discretizing the infinite-state transition probability matrix of the chain.

On the other hand, when parameters are estimated then the ARL is a random variable which is conditioned to the estimates obtained from the Phase I, and therefore is called *conditional* ARL. However, this quantity reflects only the performance for an specific chart, which does not provide a lot of information about the chart performance. When the ARL of the chart is calculated by averaging over all the possible values of the parameters, then it is called marginal or unconditional ARL. In addition to this measure, the Standard Deviation of the RL, ($SDRL$) was usually also reported in several works that evaluated control charts performance as well as some quantiles of the RL's distribution in order to have a better understanding of the chart performance properties.

Several works were devoted to compare estimators for the mean and/or the variance and to calculate the marginal ARL, $SDRL$, and other RL quantiles. These works can be found in the literature reviews of Jensen et al. (2006) and Psarakis et al. (2014), where the general conclusions include the fact that charts performance under estimated parameters is no longer the same as in the known parameters case and that a large amounts of Phase I data are required in order to diminish this difference, among others. In order to avoid this Phase I problem other approaches have been proposed, such as the self-starting methodology introduced by Hawkins (1987) and the Q-charts developed by Quesenberry (1991a). The main advantage of these methods is that they avoid the necessity of high amounts of Phase I data as the parameters are updated as soon as observations are available. For the sake of brevity, only key works are mentioned here and the reader is referred to the literature reviews mentioned above.

There are two recent works in SPM that have taken the attention of several authors due to their relation to the conditional control chart performance under parameter estimation. First, as the marginal or unconditional ARL is a random variable it also has a variation which could be though as the ARL's sampling variation or as the "practitioner-to-practitioner" variation from a manufacturing point of view. In the work of Jones and Steiner (2012) the Standard Deviation of the ARL, ($SDARL$) was introduced as a performance measure to evaluate the risk-adjusted CUSUM control charts. The rationale of these authors for considering it as a performance measure was that in manufacturing processes the monitored data tend to be homogeneous whereas in patient surveillance, they tend to be heterogeneous and therefore, it should be a difference between patients. The other work is due to Gandy and Kvaløy (2013) where a methodology to adjust control charts limits in order to guarantee an in-control performance with certain probability conditioned on the Phase I estimates based on bootstrapping is provided. It has to be noted that even though the adjustment of the control limits to have a guaranteed performance was previously considered by Albers and Kallenberg (2005) it was with Gandy and Kvaløy's work that this topic regained the attention of SPM researchers. These two works seem to provoke an increase in SPM research consisting mainly in the (re)evaluation of control charts under parameter estimation with the $AARL$ and $SDARL$ as performance measures, and the design of control charts to have a guaranteed performance, usually the in-control performance, via exact or bootstrap-based methods.

Concerning the estimated parameter issue when designing and implementing control charts, the topics of conditional and guaranteed performance were considered by SPM researchers. Most of the works have been focused on the evaluation of the control charts performance by means of $AARL$ and $SDARL$ and/or on the adjustment of the control limits. For instance, concerning attribute data Zhang et al. (2013) evaluated the geometric control charts (g-charts) whereas Lee et al. (2013), the upper-sided Bernoulli CUSUM charts. Zhao and Driscoll (2016) and Faraz et al. (2017) evaluated and adjusted the control chart limits via the bootstrap method proposed by Gandy and Kvaløy for the c and the np-chart.

Concerning the monitoring of the process mean, Zhang et al. (2014) evaluated the Exponential CUSUM control chart considering both the marginal and conditional ARL distribution. They also suggested to have an $SDARL$ within 10% the nominal ARL value. Saleh et al. (2015b) evaluated the performance of the $\bar{X}$ and X control charts by considering several variance estimators. For the EWMA control chart Saleh et al. (2015a) reevaluated the performance considering the $SDARL$ and adjusted the control chart limits; Aly et al. (2015b) compared the Shewhart, EWMA and Adaptive EWMA (AEWMA) control charts by means of $SDARL$ considering both the adjusted and unadjusted control limits cases; Aly et al. (2016) evaluated the Multivariate version of the AEWMA finding that around 90% of the in-control ARL distribution lies under the nominal ARL value for small Phase I sample sizes m and adjusted the MAEWMA control limits using the bootstrapping methodology proposed by Gandy and Kvaløy (2013) in order to overcome that situation. Concerning CUSUM control charts, Saleh et al. (2016) used the Markov Chain approach of Brook and Evans (1972) and the Gaussian Quadrature method to compute the integral-equations related to the computations of the $AARL$ and $SDARL$ involved there; Hany and Mahmoud (2016) evaluated the Crosier's CUSUM control chart; Jeske (2016a) and Jeske (2016b) provided the Phase I sample size required to achieve a conditional ARL_0 closer, within certain relative error, of the CUSUM control chart for exponentially distributed data and normally distributed data, respectively. The $\tilde{X}$ control chart was considered by Hu and Castagliola (2017). Finally, Goedhart et al. (2017b) developed an exact method to adjust the Shewhart $\bar{X}$ and X control chart limits.

Concerning the variance monitoring, Epprecht et al. (2015) considered the conditional distribution of the false-alarm probability, (FAP, related to the in-control ARL) of the S and S^2 control charts and provided the minimum number of Phase I data needed to guarantee, with a certain probability, that the FAP will not exceed a nominal FAP value. The study of the S^2 control chart performance in terms of the $SDARL$ was done by Faraz et al. (2015) who also used a bootstrapping methodology to adjust the chart control limits in order to lower the $SDARL$ values. The adjustment of the control charts via exact methods was done by Goedhart et al. (2017a) for the one-sided Shewhart S control chart, by Diko et al. (2017) for the R and S two-sided control charts, and by Guo and Wang (2017) for the two-sided S^2 control chart. Recently, Faraz et al. (2018) proposed exact methods to adjust the control limits of the Shewhart $\bar{X}$ and S^2 control charts to have a guaranteed conditional in-control performance. On the other hand, Aparisi et al. (2018) provided the first work to consider the trade-off between the conditional in-control and out-of-control performances for the S^2 control chart under estimated parameters. They developed a technique to determine the minimum Phase I sample size to guarantee, with a certain probability, both in-control and out-of-control performances.

Considering the self-starting methodology, Keefe et al. (2015) evaluated the performance of the Shewhart Q and CUSUM Q-charts in terms of the $SDARL$. They found that even when consistent estimators are used, the reduction on the ARL's variability is not monotonically related to the increase in the sample size, at least upon a certain size. For simple linear profiles monitoring, Aly et al. (2015a) studied the performance of three different approaches where large amounts of Phase I data are needed to have an $SDARL$ within 10% the nominal value. Cheng et al. (2018) evaluated the effect of parameter estimation on the Phase II synthetic exponential control charts for monitoring the time between events.

It is noteworthy to say that the adjustment control limits lead to high in-control ARL values, as it is often used to ensure that 90% of the charts will be over the nominal ARL value. However, this adjustment leads to an in-control ARL distribution with higher variability when compared to the case without the adjustment. Taking this in mind, Zwetsloot and Woodall (2017) proposed a definition of what should be understood as "better performance" when comparing control charts. In the proposed definition, they stated that if the difference of the ARL's obtained with each chart lies within the 5% of the nominal ARL value, those charts are considered as having an equivalent performance. Otherwise, the chart whose ARL is closer to the desired ARL is considered to have *better performance* than the other. With this definition, they made a head-to-head comparison between Shewhart, CUSUM and EWMA control charts for monitoring the process mean under parameter estimation, concluding that the EWMA and CUSUM do not have equivalent performance in opposition to the findings of Hawkins and Wu (2014) for the known parameter case.

Table 2.2.1 stands for the works done in parameter estimation by considering mainly the works that are not included in Jensen et al. (2006) and Psarakis et al. (2014) reviews and the works devoted to the conditional in-control performance and methods for adjustment of the control charts limits to guarantee conditional performance.

Table 2.2.1: Summary of works related to the i.i.d. case.

Author/Year	Contribution
Brook and Evans (1972)	Develop a methodology to compute the ARL of CUSUM control charts using a Markov chain approach.
Crowder (1987)	Develop an integral-equation method to compute the ARL of the EWMA control chart.
Hawkins (1987)	Proposed the self-starting methodology to address the Phase I problem.
Lucas and Saccucci (1990)	Modified the Markov Chain approach of Brook and Evans (1972) to compute the ARL of the EWMA control chart.
Quesenberry (1991a)	Proposed the Q-charts for known and unknown parameters.
Albers and Kallenberg (2005)	Proposed a methodology to adjust control chart limits in order to guarantee an in-control performance with certain probability.
Jensen et al. (2006)	Literature review for the effect of parameter estimation on control charts performance.
Jones and Steiner (2012)	Considered the $SDARL$ as a performance measure for control charts with estimated parameters.
Gandy and Kvaløy (2013)	Bootstrapping methodology to adjust chart control limits to guarantee an in-control performance, with certain probability, for control charts with estimated parameters.
Lee et al. (2013)	Compare the conditional in-control performance of g-charts and the Upper-sided Bernoulli CUSUM when monitoring non-conforming proportions.
Zhang et al. (2013)	Evaluate the performance of the g-charts using $SDARL$ as performance measure.
Hawkins and Wu (2014)	Head-to-head comparison of power detection of EWMA and CUSUM charts under known parameters.
Psarakis et al. (2014)	Extension of the literature review of Jensen et al. (2006).
Zhang et al. (2014)	Evaluated the Exponential CUSUM through the conditional and marginal ARL's distribution.
Aly et al. (2015a)	Compared three different approaches for simple linear profile monitoring using $SDARL$ as performance measure.

Table 2.2.1: (cont.) Summary of works related to the i.i.d. case.

Author/Year	Contribution
Aly et al. (2015b)	Compared the AEWMA, EWMA and Shewhart $\bar{X}$ control charts under parameter estimation using $SDARL$ as performance measure.
Epprecht et al. (2015)	Calculated the required Phase I sample sizes for the S and S^2 charts needed to guarantee that the false-alarm probability will not exceed some nominal value.
Faraz et al. (2015)	Adjusted S^2 chart control limits using the bootstrap method to overcome the Phase I problem.
Keefe et al. (2015)	Evaluated the conditional in-control performance of the self-starting control charts and its relation with the over-understimation of the mean and variance.
Saleh et al. (2015a)	Reevaluated the EWMA control charts with estimated parameters, considering several estimators for the process standard deviation.
Saleh et al. (2015b)	Shown that unrealistic amounts of Phase I data are required to guarantee a conditional in-control performance of $\bar{X}$ and X Shewhart charts.
Aly et al. (2016)	Evaluated the in-control performance of the MEWMA for normally distributed random vectors.
Hany and Mahmoud (2016)	Evaluated the Crosier's CUSUM control chart under estimated parameters.
Jeske (2016a)	Obtained the reference sample size needed to guarantee an in-control performance for CUSUM control chart for exponential random variables.
Jeske (2016b)	Calculate the minimum Phase I sample size to have a desired in-control performance for CUSUM control chart for normal variables.
Saleh et al. (2016)	Reevaluated the one and two-sided CUSUM control chart considering the $SDARL$.
Zhao and Driscoll (2016)	Evaluated the conditional in-control performance of the c-chart and adjusted the control limits using the bootstrap method.
Diko et al. (2017)	Proposed a modification of the Goedhart et al. (2017a) method to adjust control limits for the two sided R and S control charts.
Faraz et al. (2017)	Evaluated the conditional in-control performance of the np-chart and adjusted the control limits using the bootstrap method.
Goedhart et al. (2017a)	Developed an exact method to adjust the one-sided Shewhart S control chart to guarantee a conditional in-control performance.
Goedhart et al. (2017b)	Proposed a method to adjust control limits of the $\bar{X}$ control chart to guarantee an in-control performance. This adjusted values depends on the subgroup size (n) and the number of Phase I samples (m).
Guo and Wang (2017)	Proposed an exact method to adjust the two-sided S^2 control chart to guarantee a conditional in-control performance.
Hu and Castagliola (2017)	Evaluated the in-control performance of the median $\tilde{X}$ control chart and adjusted the control limits by a bootstrap method.
Zwetsloot and Woodall (2017)	Provide a definition of "better performance" to compare charts. Also, found that CUSUM and EWMA charts do not have an equivalent in-control performance under this definition.

Table 2.2.1: (cont.) Summary of works related to the i.i.d. case.

Author/Year	Contribution
Aparisi et al. (2018)	Developed an exact technique to determine the minimum Phase I sample size to guarantee both in-control and out-of-control performances for the S^2 control chart.
Cheng et al. (2018)	Evaluated the conditional in-control performance of a synthetic exponential control chart for monitoring the time between events.
Faraz et al. (2018)	Proposed exact methods to adjust control limits for the Shewhart $\bar{X}$ and S^2 control charts.

2.3 Self-starting control charts

It is well known that control charts performance are negatively affected when using estimations as the process true parameters values, and that it is due to the estimation error. That error could be diminished (and the chart performance improved) by considering a reference sample large enough to provide accurate estimates. This is the so-called Phase I problem under the SPM label which has been deeply studied for the independent case. Nevertheless, when the process has no historical data or when the sampling costs are quite expensive, this approach is not suitable and other approaches are needed.

This necessity of a large in-control initial sample size was avoided by Hawkins (1987) with the self-starting methodology. The idea is collect data (assumed to be independent and normally distributed random variables) and update the parameter estimations as more observations are available, constructing a sequence of independently random variables distributed as t with certain (different) degrees of freedom. Thereafter, in order to all of them have the same distribution, these statistics are again transformed by means of the t distribution and the inverse of the normal standard distribution, to generate approximately normal standard random variables. To assess the same problem, Quesenberry (1991a) proposed the Q charts considering i.i.d. normal observations using Q statistics which were developed considering process parameters as unknown and were proved to constitute a sequence of independent normal standard random variables. After that, he proposed the Q charts for data following a binomial model Quesenberry (1991b), and Q charts for attribute data from a Poisson approach Quesenberry (1992). Following the guidelines of Quesenberry (1991a,b), Zantek (2006) evaluated the performance of the CUSUM Q chart and noticed that optimal values proposed for the CUSUM chart are not the same for the CUSUM Q charts, due to there was not consideration on the chart statistic distribution under a shift in the mean. Later, considering the work of Zantek (2006) and the HQK methodology (Hawkins et al., 2003), Li and Wang (2010) studied the impact of the choice of the parameters and the masked effect on the CUSUM Q chart and designed the Adaptive CUSUM Q chart (ACQ chart) by proposing an adaptive technique to choose the chart parameters.

In order to create self-starting control charts, Hawkins et al. (2003) proposed a methodology to design this type of charts (is also known as HQK) considering several scenarios including all parameters unknown and using a change-point formulation. According to each scenario, they recursively used the test statistic in order to calculate the control limits of the chart. Hawkins and Zamba (2005) created a control chart for monitoring the variance of normally distributed processes using the GLR and the Bartlett factor correction for the equal variances test. Considering the LRT approach and the EWMA procedure, Li et al. (2010) proposed a self-starting control chart (SSELR) to monitor the mean and variance of normally distributed processes and used a two-dimensional Markov chain model to calculate the in-control ARL of the chart. Even though its advantages over other approaches, self-starting methodology produces biased charts as it was proved by He et al. (2008) for the Shewhart Q charts. However, they also proposed a modification to these charts which diminishes the bias of the charts and thus, improves the Q charts detection capability.

Concerning nonparametric approaches, the HQK methodology was considered by Zhou et al. (2009): they developed a nonparametric control chart for shifts in the mean without prior knowledge of the underlying distribution only assuming that there are m historical samples available. The chart is based on the Mann-Whitney test and it provides also a comparison with the HQK chart for a few distributions such as normal, lognormal, χ^2 and t. Based on the same test and the HQK methodology, Hawkins and Deng (2010) developed the corresponding nonparametric control chart using the change-point formulation and compared it with Zhou et al.'s chart concluding that the assumption made by the latter ones only provides benefits for small shifts. After that, Liu et al. (2013) designed a nonparametric control chart to detect shifts in the mean of continuous distributions based on sequential ranks and considering the Adaptive EWMA (AEWMA) chart proposed by Capizzi and Masarotto (2003). This chart is distribution free, which means that the ARL does not depend on the data's distribution. Recently, Liu et al. (2015) modified the AEWMA (or NAE) control chart proposed by Liu et al. (2013) using the VSI procedure.

For multivariate processes, Sullivan and Jones-Farmer (2002) extended the methodology used on the Multivariate EWMA chart (MEWMA chart) but considering that mean vector and covariance matrix as unknown to develop the SSMEWMA to monitor the mean vector. They used a low value of the false alarm probability (FAP) in order to avoid false alarms due to thicker control limits. After that, Maboudou-Tchao and Hawkins (2011) designed a control chart to detect shifts in the covariance matrix extending the Multivariate Exponentially Moving Covariance Matrix chart (MEWMC) using the self-starting methodology and the SSMEWMA to monitor both covariance matrix and mean vector, respectively. This procedure is called Self-Starting Multivariate Exponentially Weighted Moving Average and Moving Covariance Matrix (SSMEWMAC) chart.

Considering different models, Zou et al. (2007) proposed a self-starting control chart for monitoring the slope, intercept and standard deviation of linear profiles assuming unknown parameters. This chart is based on recursive residuals and two EWMA charts with the control limits calculated using the Markov Chain approach. Zhang et al. (2012) designed a self-starting cumulative count-of-conforming chart $CCCg$) using the self-starting procedure and integrating it with an approximated ARL. Recently, Capizzi and Masarotto (2012) suggested a new self-starting control chart using a procedure based on Adaptive CUSCORE, (ACUSCORE) to detect shifts in the mean. This chart uses Q statistics whose reference values are updated by means of an adaptive EWMA.

The work of Keefe et al. (2015) was mentioned in the previous section since it is devoted to the conditional in-control performance of the Shewhart Q and CUSUM Q control charts. The work of Snoussi et al. (2005) will be mention again in the next section, as it is related to control charts for autocorrelated data and seems to be better consider it there. These works are mentioned here just for the sake of completeness. As it can be seen in Table 2.3.1, almost all self-starting charts were developed for independent data, letting a gap when independence assumption is not met.

Table 2.3.1: Summary of works related to self-starting methods.

Author/Year	Contribution	Type of data.
Hawkins (1987)	Set the theoretical basis to develop self-starting charts.	Variable.
Quesenberry (1991a)	Developed Q charts for known and unknown parameters under normality.	Variable.
Quesenberry (1991b)	Develop the Q charts for binomial distributions.	Variable.
Quesenberry (1992)	Q charts for Poisson processes	Attribute.
Sullivan and Jones-Farmer (2002)	Self-starting Multivariate EWMA (SSMEWMA) for monitoring the mean vector of multivariate normal processes.	Variable.
Hawkins et al. (2003)	Methodology to create self-starting charts based on change-point model.	Variable.

Table 2.3.1: (cont.) Summary of works related to self-starting methods.

Author/Year	Contribution	Type of data.
Hawkins and Zamba (2005)	Control chart based on the LRT of a variance test with the Bartlett correction factor and using Change-point approach.	Variable.
Snoussi et al. (2005)	Evaluated the performance of the Shewhart Q and EWMA Q control charts applied to the residuals of an AR(1) process assuming parameters as known.	Variable.
Zantek (2006)	Provide optimal values for the CUSUM charts applied to Q statistics.	Variable.
Zou et al. (2007)	Self-starting chart for linear profiles.	Variable.
He et al. (2008)	Proved that Q charts are biased and proposed an improvement.	Variable.
Zhou et al. (2009)	Nonparametric control chart using the HQK methodology.	Variable.
Hawkins and Deng (2010)	Nonparametric chart for monitoring the mean based on the HQK methodology.	Variable.
Li and Wang (2010)	Propose an Adaptive CUSUM for Q statistics.	Variable.
Li et al. (2010)	Develop a control chart based on the LRT and EWMA methodology for normal processes to monitor both mean and variance.	Variable.
Maboudou-Tchao and Hawkins (2011)	Self-starting chart for multivariate normal processes for monitoring the covariance matrix and the mean vector.	Variable.
Capizzi and Masarotto (2012)	Propose a technique to develop self-starting control charts: Adaptive CUSCORE (ACUSCORE) to assess charts biasedness.	Variable.
Zhang et al. (2012)	Self-starting Cumulative count-of-conforming chart ($CCCg$).	Attribute.
Liu et al. (2013)	Nonparametric chart based on the Adaptive EWMA of Capizzi and Masarotto (2003).	Variable.
Liu et al. (2015)	Nonparametric chart for mean based on the AEWMA of Liu et al. (2013)	Variable.
Keefe et al. (2015)	Evaluated the conditional in-control performance of the self-starting control charts and its relation with the over-understimation of the mean and variance.	Variable.

2.4 SPC for autocorrelated data

Another important issue to consider when developing a control chart is the independence assumption of the data, since control charts are sensitive to departures of independence. Considering that it could not be validated nor to be suitable for monitoring processes such as chemical ones (Montgomery, 2007 contains several examples), the main efforts go from the evaluation of the control charts performance under autocorrelation for Shewhart, CUSUM and EWMA control charts to the design of control charts based on ARMA time-series models. It is noteworthy to say that the monitoring of correlated processes is more difficult than in the i.i.d. case as a model is fitted and its parameters, estimated. This increases the uncertainty on the model and on the estimations. However, these kind of processes arise in real life applications and some approaches were developed to overcome these issues. For an in-depth study of the SPC procedures to handle autocorrelated processes, the reader is referred to the literature reviews of Psarakis and Papaleonida (2007) and Prajapati and Singh (2012) and for the sake of brevity of the document, only key works in this area and recent developments will be considered here.

There are some approaches which deal with techniques to make feasible the use of traditional control charts. One of these approaches is to remove the autocorrelation of the data. This could be done by omitting certain number of collected data points or changing the sampling scheme (for instance, the variable sampling interval introduced by Reynolds et al. (1988) or the use of batch means Runger and Willemain (1996)). The main drawback of these approaches might be the fact that not all available information is used to understand the process behavior, which could be useful to understand the process dynamics. However, the effect of autocorrelation on traditional charts performance was evaluated for Shewhart (Stamboulis, 1971 for $\bar{X}$ and Amin et al., 1997 for R and S^2 charts), CUSUM (Johnson and Bagshaw, 1974) and EWMA (Harris and Ross, 1991) charts in terms of the RL distribution and it was found that the rate of false alarms increases in presence of autocorrelation, specially for high levels of autocorrelation. As control limits are constructed taking into account the process variance Stamboulis (1971) showed that AR(1) process variance differs from the process variance on the i.i.d. case and proposed to adjust control limits using the process true variance. However, this adjustment lead to the widening (or shrinking) of the control limits, which will affect both in-control and out-of-control ARL's, that is why the adjustment should be made carefully as in Vasilopoulos and Stamboulis (1978) in order to avoid a high rate of false-alarms or missing noticeable changes, as shown in Section 1.2.

Therefore, in order to monitor and detect efficiently changes in autocorrelated process, several control charts have been developed, considering mainly time series models. For instance, Alwan and Roberts (1988) proposed the Special Cause Chart (SCC or *residual* chart) which uses the residuals (instead of observed data) obtained from a fitted time series model (usually a particular ARIMA model) to collected data. However, Longnecker and Ryan (1991) show that performance of traditional control charts applied to residuals from a fitted AR(1) process is not the same that the one under the independence assumption. Aside the *residuals* control charts, some control charts for stationary processes were developed, namely the Exponentially Weighted Moving Average for Stationary Processes, EWMAST (introduced by Zhang, 1998); the Autoregressive and Moving Average for Stationary Processes, ARMAST (proposed by Jiang et al., 2000 as an extension of the EWMAST); and the distribution free tabular CUSUM, (DFTC) for autocorrelated data developed by Kim et al. (2007).

There have been other works on this line that have not been included in the literature reviews mentioned above. Among these, Croux et al. (2011) considered robust estimators used with control charts for time series data, particularly considering non-stationary time series whereas Chang and Wu (2011) proposed a Markov Chain-based approach to compute the ARL for the Shewhart, CUSUM and EWMA control charts, while Lwin (2011) considered the estimation of the autoregressive parameter ϕ_0 by using a parametric and semi-parametric approach and their implementation with the EWMAST control chart. The problem of the model selection was considered by Ledolter and Bisgaard (2011) who presented an example of real data that could be modeled sufficiently well by different time series models. The change-point analysis was considered by Wu (2016) who provided more accurate confidence intervals for the change-point location using a CUSUM control chart for AR(1) processes while De Ketelaere et al. (2016) evaluated different Principal Component Analysis (PCA)-based methods to monitor AR(1) and ARI(1,1) processes and consider the extension to multivariate time-series. Lee and Wei (2017) considered the Likelihood Ratio Test (LRT) of a change in the mean of ARMA models and they compared its power when using non-aggregated and aggregated data.

Concerning the *residuals* control charts introduced by Alwan and Roberts (1988), Snoussi et al. (2005) consider the Shewhart Q and EWMA Q control charts applied to the residuals of AR(1) processes *but* assuming that process parameters are perfectly accurate. Trying to improve the performance of these charts, Triantafyllopoulos and Bersimis (2016) proposed a modification based on a Bayes factor to detect departures from in-control situations. The problem of the effect of the model accuracy on these charts was addressed by Zhou and Goh (2016) who evaluated such effect on multivariate autoregressive process of order 1, MAR(1). Recently, Dawod et al. (2017) evaluated the performance of the residuals Shewhart, EWMA and CUSUM charts for AR(1), MA(1) and ARMA(1,1), *but* considering that process parameters estimations were sufficiently accurate.

Asides *residuals* control charts, Alshraideh and Khatatbeh (2014) considered the Gaussian Process control chart; Zhang and Pintar (2015) extended the Exponentially Weighted Mean Square, EWMS for variance monitoring; Harris et al. (2016) proposed a multivariate control chart to monitor an autocorrelated tool wear process; Dasdemir et al. (2016) evaluated the effect of two Phase I approaches when dealing with outliers on the Phase II performance of the modified AR(1) Shewhart chart. Osei-Aning et al. (2017b) provided the optimal scheme for the CUSUM and EWMA control charts for monitoring the mean of stationary AR(1) processes whereas Osei-Aning et al. (2017a) proposed the mixed EWMA-CUSUM and mixed CUSUM-EWMA control charts for the same end. Nevertheless, these works did not consider the issue of estimated parameters. Recently, Weiß et al. (2018) considered the guaranteed conditional in-control performance of the Shewhart X control chart for AR(1) processes. The adjustment of the control limits is done via both parametric and non-parametric bootstraps.

Even though the time series approach allows to deal with autoregressive and integrated moving averages ARIMA models it often requires a model selection and parameter estimation, which makes the procedure more complex than in the i.i.d. case. A summarize of the works done when dealing with autocorrelation are found in Table 2.4.1.

Table 2.4.1: Summary of works related to SPC methods for autocorrelated processes.

Author/Year	Contribution
Vasilopoulos and Stamboulis (1978)	Developed the modified $\bar{X}$ control chart for AR(1) processes.
Amin et al. (1997)	Developed the modified R and S^2 control charts for AR(1) processes and provided a way to compute the ARL of the modified S^2 control chart.
Zhang (1998)	Introduced the EWMAST control chart: an EWMA control chart for stationary processes.
Jiang et al. (2000)	Developed the ARMAST control chart: an ARMA chart for stationary processes with the EWMAST as a special case.
Snoussi et al. (2005)	Evaluated the performance of the Shewhart Q and EWMA Q control charts applied to the residuals of an AR(1) process assuming parameters as known.
Capizzi and Masarotto (2007)	Studied the effect of parameters estimation on the EWMAST control chart.
Kim et al. (2007)	Developed a distribution free tabular CUSUM for autocorrelated data.
Psarakis and Papaleonida (2007)	Literature review for SPC techniques for autocorrelated processes.
Chang and Wu (2011)	Markov Chain approach to compute the ARL for Shewhart, CUSUM and EWMA control charts for autocorrelated data.
Croux et al. (2011)	Developed a control chart for monitoring non-stationary time series under parameter estimation.
Ledolter and Bisgaard (2011)	Expose the difficulty of time-series modeling problem consider data sets that could be equally fitted to different models.
Lwin (2011)	Consider the problem of robust estimation of the autoregressive parameter of stationary AR(1) processes.

Table 2.4.1: (cont.) Summary of works related to SPC methods for autocorrelated processes.

Author/Year	Contribution
Prajapati and Singh (2012)	Summary of control charts for autocorrelated processes.
Alshraideh and Khatatbeh (2014)	Introduced the Gaussian Process control chart to monitor stationary processes, based on the multivariate normal distribution.
Zhang and Pintar (2015)	Extended the EWMS control chart for stationary processes and evaluate their performance by means of ARL and MRL.
Dasdemir et al. (2016)	Consider the effect of outliers in the performance of control charts for AR(1) processes.
De Ketelaere et al. (2016)	Review of Principal Component Analysis (PCA)-based methods to monitor autoregressive processes.
Harris et al. (2016)	A Multivariate control chart for monitoring tool wear processes.
Triantafyllopoulos and Bersimis (2016)	Propose a method based on a Bayesian factor to detect departures from an in-control time series model.
Wu (2016)	Provides bias and pivots for the change-point considering an AR(1) process.
Zhou and Goh (2016)	Evaluated the effect of model accuracy on the performance of residuals charts for multivariate AR(1) processes.
Dawod et al. (2017)	Compared the Shewhart, CUSUM and EWMA residuals control chart for AR(1), MA(1) and ARMA(1,1) processes.
Lee and Wei (2017)	Studied the effect of aggregation on the LRT for changes in the mean of stationary AR(1) processes.
Osei-Aning et al. (2017b)	Provide the optimal EWMA and CUSUM schemes to be used for stationary AR(1) processes.
Osei-Aning et al. (2017a)	Proposed mixed EWMA-CUSUM and mixed CUSUM-EWMA control charts for monitoring AR(1) processes.
Weiß et al. (2018)	Bootstrapping methods to adjust the Shewhart X control chart for AR(1) processes.

2.5 Gap Analysis

It has been proved that control charts are helpful for assist managers on the decision-making process as they are good to detect special causes of variation. However, their efficiency is negatively affected when the assumptions are not met by the process, situation that arises naturally in several real life applications, such as the parameters estimation issue and the autocorrelated data. As it can be seen from Tables 2.2.1, 2.3.1 and 2.4.1, there has been many efforts to deal with these issues trying to provide guidelines for practitioners in the searching of assignable causes of variation.

Considering independent observations and the Phase I problem, the collection of an initial in-control sample of a sufficiently large size as a solution approach was deeply studied and considered by many authors. However, for start-up processes or those without historical data that approach is not feasible and therefore, the self-starting methodology or the Q-transformations are preferred. Even though these methodologies provide biased charts, some authors have been proposed modifications in order to improve Self-Starting Control Charts (SSCC) performance. Furthermore, the performance of this kind of charts is suggested to be done in terms of the $AARL$ and the $SDARL$, since there have a variation due to the initial-sample used to start the recursion for updating the parameters/statistics.

When dealing with autocorrelation, residuals charts were proposed as a solution approach by considering time series model which allows to use all available data at hand and also to avoid those approaches which are not suitable when considering short runs, such as skipping several observations to diminish the autocorrelation. Nevertheless, as in the independent case, there is a parameter estimation issue which has not been deeply studied yet from a self-starting point of view and/or considering the $SDARL$ as a performance measure for the autocorrelated case. There are some studies about the effect of parameters estimation when dealing with autocorrelated data (e.g. Capizzi and Masarotto (2007) for the EWMAST chart), but no one considering the $SDARL$ as performance measure was found.

Summarizing these facts and all previous comments, we have that:

- There are several works considering the effect of parameter estimation on the performance of traditional control charts, i.e. Shewhart, EWMA and CUSUM charts. Nevertheless, there is a lack of the effect of parameter estimation on the performance of control charts for monitoring the mean and variance of autocorrelated data. Particularly, for the modified $\bar{X}$ and S^2 control charts, which are widely used in practice under the i.i.d. case.
- There is a lack of studies about guaranteeing the conditional in-control performance of control charts for autocorrelated processes. Just one work was found, and it is quite recent.
- Concerning the $SDARL$ as a performance measure, there is a lack of studies considering a comparison of control charts for autocorrelated processes when applied to the process observations (or to the residuals).
- There is not a single work about the guaranteed conditional in-control performance based on exact methods instead of bootstrap methods.
- Only two works are mentioned on McCracken and Chakraborti (2013) about to the joint monitoring of the mean and variance in presence of autocorrelation and none of them provide a detailed study of the effect of parameter estimation on the chart performance.
- The self-starting methodology has not been applied to the residuals of fitted time series model under estimated parameters.
- When evaluating control charts under estimated parameters there are two approaches: (i) try to get an small $SDARL$ value in order to ensure that chart's performance is near to some nominal ARL value, and (ii) have a guaranteed conditional in-control performance. The latter one usually is reached by adjusting (usually widening) the control limits, leading to higher ARL values that inflate the $SDARL$. According to the approach some charts might be preferred over others or even over their "adjusted versions". There is a lack of head-to-head comparisons of control charts for autocorrelated processes: applied to the observations or to the residuals.
- The problem of model selection has not been deeply studied for control charts for autocorrelated data.

Therefore, there are several lines to follow, but the line to follow up in this research is first to evaluate the performance of the modified Shewhart $\bar{X}$ and S^2 control charts to try to see if these tools are good enough to be used in practice, since they are widely used in the i.i.d. case, and if it is possible to have a guaranteed performance with certain probability, just as the actual tendency on the i.i.d. case. After that, the joint monitoring of the mean and variance could be addressed. Finally, *residuals* control charts might be compared to these tools developed for observations in order to see if there is any advantage in using residuals as they are independent as long as parameters are well estimated and the model well specified.

As it is not possible to deal with all of these research lines in this document only the first part is considered here: the evaluation of the modified $\bar{X}$ and S^2 control charts. Therefore, the effect of autocorrelation estimators on the conditional in-control performance of the $\bar{X}$ chart for stationary AR(1) processes is detailed in Chapter 3 whereas in Chapter 4 the performance of the $\bar{X}$ control chart considering all parameters as unknown is studied. In Chapter 5, the modified S^2 chart when the process variance is estimated is evaluated.

Chapter 3. Effect of autocorrelation estimators on the performance of the $\bar{X}$ control chart

Abstract

Control charts are powerful Statistical Process Monitoring tools to detect departures from in-control situations. However, their power detection relies on the fact that all assumptions underlying their design are met, such as independence of data and knowledge of the process model parameters. When parameters are estimated, the Average and the Standard Deviation of the ARL, ($AARL$ and $SDARL$, respectively) are used as performance measures as they summarize the variation due to the Phase I estimations. Considering these performance measures, the effect of several autocorrelation estimators on the $\bar{X}$ chart performance was investigated in case of stationary $AR(1)$ processes. Further, a bootstrapping technique was developed to adjust the corresponding control limits and obtain a guaranteed ARL performance. The effect on the out-of-control ARL due to this adjustment is also presented. Results show that overestimation of the autoregressive parameter leads to higher values of both in-control and out-of-control ARL's.

Keywords: AR(1) process; $\bar{X}$ control chart; estimated parameters; Standard Deviation of the ARL; guaranteed performance.

3.1 Introduction

Statistical Process Monitoring (SPM) is a collection of statistical techniques used to assess if a process is in statistical control or special causes of variation are present and need to be addressed. Control charts are one of the most known and widely used SPM tools due to their ability to detect shifts from a target value. However, to correctly implement a control chart several assumptions should be made, such as defining the distribution of the observed data, establishing prior knowledge of the parameters, considering the independence of observations, among others. The correct estimation of the control chart's performance strongly relies on the satisfaction of these assumptions whereas, in practice, they are seldom met.

Autocorrelated streams of data arise when taking measurements of the same object, when considering a continuous flow of data, or when data are collected within small periods of time. Practical examples are found in health surveillance, crop monitoring, chemical processes, etc. Technological advances make possible to gather data at high rates, making the series likely to be autocorrelated. In these cases, traditional control charts should not be used carelessly, since their performance is known to degrade due to the presence of correlation (Psarakis et al., 2014). If a practitioner wants to implement these charts under these circumstances, an approach to decrease the level of correlation and improve charts performance is to collect data at a lower rate or skip some observations as in Costa and Castagliola (2011).

When the process parameters are unknown they are usually estimated from an (assumed) in-control historical data sample (called Phase I, Chakraborti et al., 2009) and, therefore, the chart is set up and used for process monitoring (Phase II). Parameter estimation allows to define control charts by using estimates as the parameters true values. This leads to a decrement in performance due to the added variability of estimations – which diminishes as more Phase I data is gathered and consistent statistics are used as estimators. Performance of control charts with estimated parameters has been deeply studied. For instance, Jones-Farmer et al. (2001), Chakraborti (2006), Bischak and Trietsch (2007), Graham et al. (2012), Teoh et al. (2014), Teoh et al. (2014) and Yeong et al. (2015) have focused on location type control charts (i.e. $\bar{X}$ and $\tilde{X}$); whereas Maravelakis et al. (2002), Castagliola et al. (2009) and Castagliola and Maravelakis (2011) have focused on dispersion type control charts (i.e. S, S^2 and R); and Braun (1999), Chakraborti and Human (2006), Castagliola and Wu (2012) and Castagliola et al. (2014) have focused on attribute type charts. For an in depth review on the effect of parameters estimation see Jensen et al. (2006) and Psarakis et al. (2014).

On the other hand, when the independence assumption is not suitable, control charts for autocorrelated processes have been investigated, often assuming a specific time series model. For instance, considering autoregressive processes of order 1, i.e. AR(1), Kramer and Schmid (2000) studied the performance of the Shewhart residual chart and the modified Shewhart chart when parameters are unknown while, Lwin (2011), proposed a parametric EWMAST approach based on variograms to adjust the control limits in order to avoid a high false alarm rate of the EWMAST chart for AR(1)-EIV processes, that is, AR(1) process with errors in the variables. For Autoregressive and Moving Average (ARMA) processes, Chin and Apley (2008) investigated the performance and the robustness of several residuals control charts and Lee and Apley (2011) proposed an improved design of the EWMA chart using a Bayesian approach for widening the control chart limits.

Recently, the performance evaluation of control charts with estimated parameters has been revisited by taking into account the variation due to the Phase I samples used to estimate the process parameters into account. A new performance indicator has been proposed, the standard deviation of the average run length ($SDARL$), also called as the “practitioner-to-practitioner” variance which is a performance measure summarizing this variation. The first ones who advocated the use of this new indicator were Jones and Steiner (2012) by considering a health surveillance application and, therefore, the heterogeneity of the population when applying control charts. After that, Gandy and Kvaløy (2013) proposed an efficient bootstrapping technique to provide a guaranteed performance with a certain probability conditioned on the Phase I estimations. Considering the ARL’s mean and standard deviation, $AARL$ and $SDARL$, respectively, Zhang et al. (2014) suggested to guarantee an $SDARL$ value within 5% to 10% of the target in-control ARL.

As a result, based on these considerations, the performance of the control charts under estimated parameters have been revisited using the $SDARL$: the Shewhart $\bar{X}$ and X control charts were studied by Saleh et al. (2015b) and Goedhart et al. (2017b), the EWMA $\bar{X}$ control chart by Saleh et al. (2015a) whereas the Adaptive EWMA (AEWMA) $\bar{X}$ chart by Aly et al. (2016) and its multivariate version (MAEWMA) by Aly et al. (2015b). The one and two-sided CUSUM control charts were revisited by Saleh et al. (2016) who used a Markov Chain approach to calculate the ARL and the Gaussian-Quadrature method to compute the $AARL$ and $SDARL$. The in-control performance of the CUSUM and EWMA control charts was studied by Hawkins and Wu (2014) under the known parameters case and by Zwetsloot and Woodall (2017) for the unknown parameters case. It was found that these charts have not a similar in-control performance under parameters estimation. The self-starting Q and CUSUM Q control charts were studied by Keefe et al. (2015) who found that the amount of variation using these charts is lower than those obtained by using traditional Phase I methods but the authors warned that process estimation with few data may lead into data contamination. Considering variance monitoring Faraz et al. (2015) evaluated the performance of the S^2 control chart whereas Epprecht et al. (2015) studied the effect on the false alarm rate of the S and S^2 charts with estimated parameters.

In this research, using the $AARL$ and $SDARL$ as performance measures, we investigate the $\bar{X}$ control chart for monitoring the mean of an AR(1) process with estimated parameters. We particularly focus on the effect on the chart performance due to the estimation of the autoregressive parameter by considering five estimators including the least squares, unbiased and robust estimators for the autoregressive parameter, and we provide adjusted control limits to ensure a desired in-control ARL (ARL_0) with a certain probability.

The remaining of the chapter is organized as follows: in Section 3.2 the $\bar{X}$ control chart for an AR(1) process is introduced; Section 3.3 presents the design of the simulation to evaluate the chart’s performance as well as the results. Section 3.4 provides the adjusted control limits to guarantee a conditional in-control performance and the effect of this adjustment on the out-of-control ARL (ARL_1) is commented. Finally, conclusions and future works are discussed in section 3.5.

3.2 Shewhart $\bar{X}$ chart for autocorrelated data

3.2.1 Known parameter case

Consider a sequence of observations $\{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\}$ following an autoregressive first order model AR(1):

$$X_{i,j} - \mu_0 - \delta\sigma_0 = \phi_0(X_{i,j-1} - \mu_0 - \delta\sigma_0) + \varepsilon_{i,j}, \qquad (3.2.1)$$

for $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, n,$ where n is the subgroup size at time i, μ_0 and σ_0 are the in-control process mean and standard deviation, respectively, $\varepsilon_{i,j} \sim N(0, \sigma_\varepsilon)$ are independent, $\phi_0 \in (-1, 1)$ is the in-control autoregressive parameter, $\delta = \frac{|\mu_0 - \mu_1|}{\sigma_0}$ is the standardized mean shift from μ_0 to μ_1 and $X_{i,0}$ is assumed to have the steady-state distribution, i.e. $X_{i,0} \sim N(\mu_0, \sigma_0)$. If $\delta = 0$ the process is considered as in-control, otherwise, it is considered as out-of-control. The process standard deviation σ_0 is related to σ_ε by means of

$$\sigma_0 = \sqrt{\frac{\sigma_\varepsilon^2}{1 - \phi_0^2}}. \qquad (3.2.2)$$

The $\bar{X}$ control chart uses the statistic $\bar{X}_i = \frac{X_{i,1} + X_{i,2} + \cdots + X_{i,n}}{n}$ for monitoring the mean μ_0 of an AR(1) process. Considering that, for $j = 1, 2, \ldots, n,$ $X_{i,j}$ and $X_{i+1,j}$ are independent, the standard deviation $\sigma(\bar{X}_i)$ is given by

$$\sigma(\bar{X}_i) = \frac{\sigma_0}{\sqrt{n}C_2} \qquad (3.2.3)$$

where

$$C_2 = \sqrt{\frac{n}{n + 2\left(\frac{\phi_0^{n+1} - n\phi_0^2 + (n-1)\phi_0}{(\phi_0 - 1)^2}\right)}} \qquad (3.2.4)$$

as long as $|\phi_0| < 1$, Alwan and Radson (1992). Note that C_2 only depends on n and ϕ_0.

When the process parameters μ_0, σ_0 (or σ_ε) and ϕ_0 are known the control limits LCL and UCL of the $\bar{X}$ chart for monitoring the process mean are given by

$$LCL = \mu_0 - K\frac{\sigma_0}{\sqrt{n}C_2}, \qquad (3.2.5)$$

$$UCL = \mu_0 + K\frac{\sigma_0}{\sqrt{n}C_2}, \qquad (3.2.6)$$

where $K > 0$ is a positive real valued constant chosen to satisfy some desired ARL_0 (for instance, $K = 3$ if we desire $ARL_0 = 370.4$ when $C_2 = 1$). The type II error β of the $\bar{X}$ chart for an AR(1) process with known parameters is given by

$$\beta = \Phi(K - \delta\sqrt{n}C_2) - \Phi(-K - \delta\sqrt{n}C_2) \qquad (3.2.7)$$

where $\Phi(\cdot)$ is the c.d.f. of the normal standard distribution. The corresponding ARL is equal to $ARL = \frac{1}{1-\beta}$. The effects of autocorrelation on the $\bar{X}$ control chart when all parameters are known have already been studied by Costa and Castagliola (2011). When some parameters are estimated, the ARL becomes a random variable as shown in the following section.

3.2.2 Estimated parameter case

In order to evaluate the effect of the estimation of the autoregressive parameter ϕ_0 on the $\bar{X}$ control chart, we will consider that the in-control process parameters μ_0 and σ_0 are known while the in-control autoregressive parameter ϕ_0 is unknown and has to be estimated from a Phase I random sample.

As we are considering autocorrelation within samples and not between samples then the Phase I is done by collecting m consecutive observations in order to capture the information contained on the autoregressive parameter of the $AR(1)$ process. With this sample of size m, ϕ_0 is estimated and the control chart is set up for Phase II, where samples of n subsequent autocorrelated observations are taken to compute the sample mean. These Phase II samples are collected sufficiently separated over time (say, every h hours) in order to the sample means could be considered as independent. This follows a common practice in the industry where consecutive items are measured as part of a sample, but samples are separated over time, leading to reduced correlation between samples to the point where autocorrelation can be ignored.

Therefore, the model for Phase I could be written as:

$$X_j - \mu_0 = \phi_0 \left(X_{j-1} - \mu_0\right) + \varepsilon_j \tag{3.2.8}$$

with $X_0 \sim N\left(\mu_0, \sigma_0\right)$. This model will be used through this paper when considering Phase I.

Letting $X_1, X_2, \ldots, X_m$ be the Phase I sample, then five estimators were considered for this research. First:

$$\hat{\phi}_{0,LS} = \frac{\sum_{j=2}^{m} X_j X_{j-1}}{\sum_{j=1}^{m-1} X_j^2}, \tag{3.2.9}$$

the least squares (LS) estimator, which have been proved to be biased by Marriot and Pope (1954) and Kendall (1954). Shenton and Johnson (1965) derived the first moment of this estimator, for small $|\phi_0|$, by the series

$$E(\hat{\phi}_{0,LS}) = \phi_0 - \frac{2\left(m-2\right)\phi_0}{\left(m+1\right)^{[2]}} + \frac{12\phi_0^3}{\left(m+5\right)^{[3]}} + \frac{18\left(m+8\right)\phi_0^5}{\left(m+9\right)^{[4]}} + \frac{24\left(m+10\right)\left(m+12\right)\phi_0^7}{\left(m+13\right)^{[5]}} + \cdots,$$

where $m^{[s]} = m\left(m-2\right)\ldots\left(m-2\left(s-1\right)\right)$, or as the asymptotic series (when m is large):

$$E(\hat{\phi}_{0,LS}) = \phi_0 - \frac{2\phi_0}{m} + \frac{4\phi_0}{m^2} - \frac{2\phi_0\left(1-8\phi_0^2+4\phi_0^4\right)}{m^3\left(1-\phi_0^2\right)^2} + \frac{4\phi_0\left(1-30\phi_0^2+12\phi_0^4-4\phi_0^6\right)}{m^4\left(1-\phi_0^2\right)^3} - \cdots.$$

From where it can be seen that the bias of $\hat{\phi}_{0,LS}$ is $O\left(1/m\right)$. Further, $E(\hat{\phi}_{0,LS})$ could be written as:

$$E(\hat{\phi}_{0,LS}) = \phi_0\left(1 - \frac{2\left(m-2\right)}{\left(m+1\right)^{[2]}}\right) + O\left(1/m^3\right)$$

or as:

$$E(\hat{\phi}_{0,LS}) = \phi_0\left(1 - \frac{2}{m} + \frac{4}{m^2}\right) + O\left(1/m^3\right)$$

which leads to the creation of two less biased estimators:

$$\hat{\phi}_{0,LS1} = \hat{\phi}_{0,LS} \cdot \left(1 - \frac{2\left(m-2\right)}{\left(m+1\right)^{[2]}}\right)^{-1} = \left(\frac{m^2-1}{m^2-2m+3}\right)\hat{\phi}_{0,LS}$$

and

$$\hat{\phi}_{0,LS2} = \hat{\phi}_{0,LS}\left(1 - \frac{2}{m} + \frac{4}{m^2}\right)^{-1} = \left(\frac{m^2}{m^2-2m+4}\right)\hat{\phi}_{0,LS} \tag{3.2.10}$$

whose bias is $O\left(1/m^3\right)$. Given the similarity between these estimators, we only consider the one shown in equation (3.2.10).

A third estimator considered in this research was proposed by Quenouille (1949). It is an unbiased, least-squares based estimator given by

$$\hat{\phi}_{0,Q} = 2\hat{\phi}_{0,LS} - \frac{1}{2}\left(\frac{\sum_{j=2}^{\lfloor m/2\rfloor} X_j X_{j-1}}{\sum_{j=1}^{\lfloor m/2\rfloor -1} X_j^2} + \frac{\sum_{j=\lfloor m/2\rfloor+2}^{m} X_j X_{j-1}}{\sum_{j=\lfloor m/2\rfloor+1}^{m-1} X_j^2}\right), \tag{3.2.11}$$

where $\lfloor\cdot\rfloor$ is the rounded down integer function. Aside from the least-squares based estimators, and also considered in this paper as our fourth estimator, Hurwicz (1950) proposed an estimator based on the median of the ratios:

$$\hat{\phi}_{0,H} = \text{median}\left(\frac{X_2}{X_1}, \frac{X_3}{X_2}, \ldots, \frac{X_m}{X_{m-1}}\right) \tag{3.2.12}$$

in order to provide robustness over outlying observations and data contamination. Finally, the fifth estimator under consideration was proposed by Haddad (2000) with the introduction of the median substitute estimator $\hat{\phi}_{0,MS}$ as the solution of the quadratic equation:

$$\text{sign}\,(\phi)\, 0.26\phi^2 + 0.195\phi - 0.4705 \times \frac{\text{median}(X_1X_2, \ldots, X_{m-1}X_m)}{\text{median}(X_1^2, X_2^2, \ldots, X_{m-1}^2)} = 0. \tag{3.2.13}$$

The distribution of the least squares estimator $\hat{\phi}_{0,LS}$ has been deeply studied before due to its relation with the invertibility and/or stationarity of AR(1) processes as well as for the development of hypotheses tests. Due to the lack of a closed form for the distribution function of $\hat{\phi}_{0,LS}$, some approximations have been suggested Wang (1992),Ali (2002), as well as other estimators, such as the Yule-Walker, Burg and a modified least squares estimator. However, the Least Squares estimator is less biased among those estimators Provost and Sangel (2005).

Therefore, when ϕ_0 is unknown and it is estimated with any of the previous estimators $\hat{\phi}_0 \in \{\hat{\phi}_{0,LS}, \hat{\phi}_{0,LS2}, \hat{\phi}_{0,Q}, \hat{\phi}_{0,H}, \hat{\phi}_{0,MS}\}$ in (3.2.9) to (3.2.13), respectively, the (estimated) control limits of the Shewhart $\bar{X}$ chart for autocorrelated data become

$$\widehat{LCL} = \mu_0 - K\frac{\sigma_0}{\sqrt{n}\hat{C}_2}, \tag{3.2.14}$$

$$\widehat{UCL} = \mu_0 + K\frac{\sigma_0}{\sqrt{n}\hat{C}_2}, \tag{3.2.15}$$

where

$$\hat{C}_2 = \sqrt{\frac{n}{n + 2\left(\frac{\hat{\phi}_0^{n+1} - n\hat{\phi}_0^2 + (n-1)\hat{\phi}_0}{(\hat{\phi}_0 - 1)^2}\right)}}. \tag{3.2.16}$$

as long as $\left|\hat{\phi}_0\right| < 1$ in order to equation (3.2.16) be valid.

3.3 Conditional performance of the $\bar{X}$ for an AR(1) process when ϕ_0 is estimated

In this section, the $\bar{X}$ control chart for an AR(1) process is evaluated considering the average and standard deviation of the ARL, i.e. $AARL$ and $SDARL$, respectively, as performance measures. There are two possible approaches to compute these measures:

- either use the p.d.f. (if known) of the ARL with a numerical method for integration, such as the Gaussian quadrature method,
- or run extensive Monte Carlo simulations to get a sufficiently precise estimation.

The second approach is chosen as there is no closed-form for the ARL distribution. The conditional type II error of the $\bar{X}$ chart for an AR(1) process is

$$\beta = \mathbb{P}(\bar{X}_i \in [\widehat{LCL}, \widehat{UCL}] | \hat{\phi}_0)$$

which could be written, after some manipulations, as:

$$\beta = \Phi\left(K\frac{C_2}{\hat{C}_2} - \delta C_2\sqrt{n}\right) - \Phi\left(-K\frac{C_2}{\hat{C}_2} - \delta C_2\sqrt{n}\right). \tag{3.3.1}$$

Given a fixed value of the estimate $\hat{\phi}_0$ obtained in Phase I, the chart statistics (the sample means $\bar{X}_i$) are assumed to be independent. Therefore, the run length of the $\bar{X}$ chart for an AR(1) process follows a geometric distribution of parameter β and, consequently, the conditional ARL given a sample is equal to

$$ARL = \frac{1}{1-\beta}$$

As a first step, the effect of the estimation of ϕ_0 on the ARL was studied by considering values of $\hat{\phi}_0 \in (-1,1)$ for selected true values of $\phi_0 \in \{\pm 0.1, \pm 0.5, \pm 0.9\}$. These values were considered in order to examine low, medium and high correlation levels for both positive and negative values. By considering ϕ_0, n and δ as fixed then the ARL is a function only of $\hat{C}_2$ or, more specifically, of $\hat{\phi}_0$.

Concerning the in-control case ($\delta = 0$) with target $ARL_0 = 370.4$, the relation between ARL and $\hat{\phi}_0$ is illustrated in Figure 3.3.1 for different values of ϕ_0. Only some values are considered here, but the curves have the same shape for any values of $\phi_0 \in (-1,1)$ and remain the same for different ARL_0 values. It can be seen from Figure 3.3.1 that it is more likely to have large ARL values when ϕ_0 is overestimated, despite ϕ_0 is positive or negative. It also can be seen that small variations from the true ϕ_0 value leads to relatively big variations on the conditional ARL.

On the other hand, concerning the out-of-control case for $\delta = 1$ (in this case, no reference line is shown since ARL_1 differs from one to another value of ϕ_0), the relation between ARL and $\hat{\phi}_0$ is illustrated in Figure 3.3.2. It can be noticed that as $\hat{\phi}_0$ increases so ARL_1 does, as in the in-control case. Thus, overestimating ϕ_0 leads to larger values of ARL than those obtained under known parameter case.

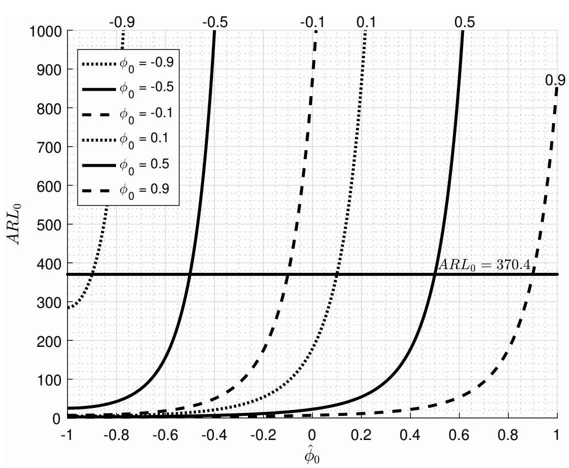


Figure 3.3.1: ARL_0 vs $\hat{\phi}_0$, for $n = 5$, $\delta = 0$, (in-control process)

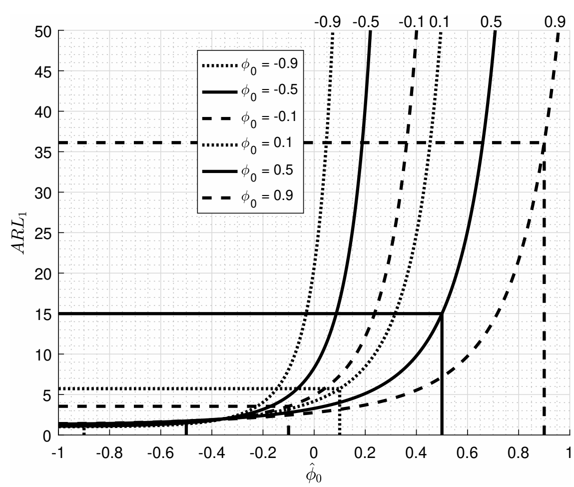


Figure 3.3.2: ARL_1 vs $\hat{\phi}_0$, for $n = 5$, $\delta = 1$, (out-of-control process)

As the values of $\hat{\phi}_0$ could vary from sample to sample (and from estimator to estimator), the conditional in-control performance of the $\bar{X}$ control chart is evaluated considering all the estimators introduced in Section 3.2.2 using Algorithm 1. The analysis was done by considering different Phase I sample sizes $m \in \{50, 100, 200, 500, 1000\}$, different values for the autoregressive parameter $\phi_0 \in \{\pm 0.1, \pm 0.5, \pm 0.9\}$ with a target $ARL_0 = 370.4$, samples of size $n = 5$, $\delta = 0$ and $rep = 10000$, where ϕ_{EST} is one of the estimators shown in equations (3.2.9) to (3.2.13).

Algorithm 1 Calculation of $AARL$, $SDARL$ and $MARL$

Define m, n, ϕ_0, ARL_0, δ, rep, ϕ_{EST}.
$K \leftarrow \Phi^{-1}\left(1 - \frac{1}{2ARL_0}\right)$.
$C_2 \leftarrow$ Calculate using equation (3.2.4) with ϕ_0.
$r \leftarrow 1$.
while $r \leq rep$ **do**
  Generate $X_1, X_2, \ldots, X_m$ based on the AR(1) model in (3.2.8)
  $X_j - \mu_0 - \delta\sigma_0 = \phi_0(X_{j-1} - \mu_0 - \delta\sigma_0) + \varepsilon_j$, where $\varepsilon_j \sim N\left(\mu_0, \sigma_0\sqrt{1-\phi_0^2}\right)$
  Estimate ϕ_0 with ϕ_{EST} and denote it as $\hat{\phi}_{0,r}$.
  if $\left|\hat{\phi}_{0,r}\right| < 1$ **then**
    $\hat{C}_{2,r} \leftarrow$ Calculate using equation (3.2.16) with $\hat{\phi}_{0,r}$ and denote it by $\hat{C}_{2,r}$.
    $\beta_r \leftarrow \Phi\left(K\frac{C_2}{\hat{C}_{2,r}} - \delta C_2\sqrt{n}\right) - \Phi\left(-K\frac{C_2}{\hat{C}_{2,r}} - \delta C_2\sqrt{n}\right)$.
    $ARL_r \leftarrow \dfrac{1}{1-\beta_r}$.
    $r \leftarrow r + 1$.
  end if
end while
$AARL \leftarrow \mathrm{mean}(ARL_1, \ldots, ARL_{rep})$.
$SDARL \leftarrow \mathrm{stdev}(ARL_1, \ldots, ARL_{rep})$.
$MARL \leftarrow \mathrm{median}(ARL_1, \ldots, ARL_{rep})$.

Tables 3.3.1 and 3.3.2 show the resulting in-control mean ($AARL_0$), standard deviation ($SDARL_0$) and median ($MARL_0$) of the ARL_0 for all estimators under study. From these measurements, we can drawn the following conclusions:

- When ϕ_0 is fixed and m increases the $SDARL_0$ decreases and the $AARL_0$ gets closer to the desired $ARL_0 = 370.4$ for both positive and negative values of ϕ_0, as well as for all ϕ_0 estimators.
- When m is fixed and ϕ_0 increases, most cases show that $AARL_0$ gets closer to the desired ARL_0 value and $SDARL_0$ decreases in almost all cases. This is not observed when ϕ_0 changes from -0.9 to -0.5. This apparent contradiction can be explained by observing Figure 3.3.1, where the slope of the ARL_0 is higher around the neighborhood of $\hat{\phi}_0$ = -0.5 than around the neighborhood of $\hat{\phi}_0$ = -0.9. The farther the slope is from zero, the bigger the variation of the ARL_0 is expected to be.
- With the constraint $\left|\hat{\phi}_0\right| < 1$, ARL values obtained when $\phi_0 = 0.9$ are bounded above by 850 whereas for $\phi_0 = -0.9$, are bounded below by 280 (see Figure $3.3.1$).
- The robust estimators $\widehat{\phi}_{0,H}$ and $\widehat{\phi}_{0,MS}$ performs poorly in almost all cases. They have quite large $SDARL_0$ values, specially for small values of m and ϕ_0.
- The value of $MARL_0$ is close to the nominal ARL_0 value for different estimators and Phase I sample sizes in most of the cases.

In order to explain the last point, note from Figure 3.3.1, that ARL_0 is a monotonically increasing function of $\hat{\phi}_0$ in $(-1, 1)$ for every ϕ_0 value. Now, as quantiles are preserved under monotonic increasing functions, therefore quantiles of ARL_0 coincides with the ARL_0 obtained with the corresponding quantile of $\hat{\phi}_0$. Taking a look at ϕ_0 estimations (which are not included here for the sake of brevity), the median of $\hat{\phi}_0$ is close to

Table 3.3.1: Descriptive statistics for the ARL_0 distribution considering $n = 5$, a desired $ARL_0 = 370.4$ and $\phi_0 \in \{0.9, 0.5, 0.1\}$

m	$\hat{\phi}_0$	$\phi_0 = 0.9$		$\phi_0 = 0.5$		$\phi_0 = 0.1$	
		$AARL_0(SDARL_0)$	$MARL_0$	$AARL_0(SDARL_0)$	$MARL_0$	$AARL_0(SDARL_0)$	$MARL_0$
50	$\widehat{\phi}_{0,LS}$	280.44 (129.86)	267.55	547.11 (670.61)	340.34	885.72 (2631.74)	366.99
	$\widehat{\phi}_{0,LS2}$	370.60 (185.58)	347.93	690.28 (952.81)	397.58	1019.87 (3680.05)	378.52
	$\widehat{\phi}_{0,Q}$	339.07 (175.15)	316.46	793.40 (1989.57)	384.72	1138.41 (4981.59)	381.89
	$\widehat{\phi}_{0,H}$	325.92 (204.47)	285.32	1824.36 (6777.25)	363.23	21324.42 (646297.67)	369.40
	$\widehat{\phi}_{0,MS}$	345.11 (226.58)	301.25	3111.37 (10151.83)	451.57	47476.01 (1.33×10^6)	374.43
100	$\widehat{\phi}_{0,LS}$	327.95 (110.85)	324.07	449.05 (340.80)	359.72	559.70 (713.48)	372.15
	$\widehat{\phi}_{0,LS2}$	378.06 (133.21)	372.27	494.66 (393.84)	389.38	581.33 (783.18)	378.08
	$\widehat{\phi}_{0,Q}$	373.69 (144.89)	358.55	504.66 (447.29)	386.76	589.81 (860.77)	377.15
	$\widehat{\phi}_{0,H}$	369.84 (179.69)	342.49	791.26 (1521.40)	373.39	1538.25 (12517.98)	370.72
	$\widehat{\phi}_{0,MS}$	392.59 (209.10)	364.79	1321.98 (4101.09)	436.41	2495.16 (41057.60)	360.42
200	$\widehat{\phi}_{0,LS}$	352.30 (84.13)	350.70	409.06 (210.76)	363.21	443.02 (288.92)	371.68
	$\widehat{\phi}_{0,LS2}$	378.43 (92.34)	376.40	427.53 (224.90)	377.90	448.49 (297.36)	374.62
	$\widehat{\phi}_{0,Q}$	379.20 (102.25)	371.74	429.85 (232.43)	377.54	449.68 (303.13)	372.96
	$\widehat{\phi}_{0,H}$	378.19 (139.58)	362.03	521.72 (519.11)	368.42	627.09 (1194.65)	372.79
	$\widehat{\phi}_{0,MS}$	428.56 (186.30)	407.22	694.44 (962.07)	427.22	691.26 (1256.27)	355.50
500	$\widehat{\phi}_{0,LS}$	364.45 (55.32)	364.41	385.40 (121.85)	367.42	399.16 (150.67)	372.20
	$\widehat{\phi}_{0,LS2}$	375.04 (57.44)	374.96	391.86 (124.84)	373.31	400.69 (152.04)	373.38
	$\widehat{\phi}_{0,Q}$	375.10 (59.69)	373.52	392.28 (126.00)	372.69	400.91 (152.85)	372.74
	$\widehat{\phi}_{0,H}$	375.09 (89.81)	368.89	417.39 (222.84)	367.46	450.48 (301.16)	373.94
	$\widehat{\phi}_{0,MS}$	447.76 (147.29)	430.20	511.06 (346.83)	416.50	447.91 (321.78)	356.41
1000	$\widehat{\phi}_{0,LS}$	368.30 (40.01)	367.79	377.60 (83.71)	368.19	382.32 (96.62)	370.20
	$\widehat{\phi}_{0,LS2}$	373.62 (40.77)	373.08	380.69 (84.70)	371.13	382.98 (97.02)	370.78
	$\widehat{\phi}_{0,Q}$	373.76 (41.62)	372.91	380.73 (85.07)	371.07	383.08 (97.19)	370.89
	$\widehat{\phi}_{0,H}$	374.13 (64.62)	370.74	395.07 (143.58)	368.26	403.78 (165.53)	371.46
	$\widehat{\phi}_{0,MS}$	445.80 (113.76)	432.92	459.61 (203.46)	417.95	394.10 (179.54)	353.07

ϕ_0 in most of the cases. This means that around 50% of the times, $\hat{\phi}_0$'s obtained are below/above ϕ_0 and since $ARL_0 = 370.4$ when $\hat{\phi}_0 = \phi_0$, then ARL_0's obtained are around 50% of the times below/above the nominal value. Thus, $MARL_0$ is close to the nominal value in several cases.

The fact that in several cases $AARL_0$ is far from the nominal value might be explained with the same argument. Despite the quantiles of the ARL_0 and $\hat{\phi}_0$ values will coincide, this is not true for the expected value due to the slope of the curve. However, that around 50% of the charts will have an ARL_0 below the nominal value is undesirable, but it could be assessed by adjusting control limits to achieve a desired value (with a certain degree of confidence). A solution of this problem is proposed in the following section with an analysis of the effect on the ARL_1.

3.4 Guaranteed conditional in-control performance control limits

The results presented in the previous section show that the ARL performance of the $\bar{X}$ control chart for an AR(1) process is around 50% of the times below the nominal ARL_0 value. In order to assess this problem and to have a guaranteed conditional in-control performance (with a certain probability), a bootstrapping methodology introduced by Gandy and Kvaløy (2013) was used. With this technique it is possible to have a minimum ARL_0 value with a certain probability by adjusting the chart's control limits.

If $\delta = 0$ and n is fixed, the $AR(1)$ process in equation (3.2.8) reduces to

$$X_j - \mu_0 = \phi_0 \left(X_{j-1} - \mu_0 \right) + \varepsilon_j$$

with $X_0 \sim N\left(\mu_0, \sigma_0\right)$ which only depends on the unknown parameter ϕ_0. Therefore, this reduced model is denoted here as $P = P(\phi_0)$. Using this notation, the proposed procedure is the following:

Table 3.3.2: Descriptive statistics for the ARL_0 distribution considering $n = 5$, a desired $ARL_0 = 370.4$ and $\phi_0 \in \{-0.1, -0.5, -0.9\}$

m	$\hat{\phi}_0$	$\phi_0 = -0.1$		$\phi_0 = -0.5$		$\phi_0 = -0.9$	
		$AARL_0(SDARL_0)$	$MARL_0$	$AARL_0(SDARL_0)$	$MARL_0$	$AARL_0(SDARL_0)$	$MARL_0$
50	$\widehat{\phi}_{0,LS}$	1063.67 (5464.59)	376.95	1853.01 (13644.16)	411.86	1179.16 (9423.53)	474.42
	$\widehat{\phi}_{0,LS2}$	1151.32 (7487.81)	365.35	1698.42 (13897.72)	343.57	881.94 (6840.31)	384.34
	$\widehat{\phi}_{0,Q}$	1444.02 (21264.86)	368.74	1754.56 (12706.19)	356.56	1038.41 (10431.19)	412.57
	$\widehat{\phi}_{0,H}$	460504.25 (3.89×10^7)	371.30	193994.11 (8.63×10^6)	379.81	16438.14 (583621.97)	448.26
	$\widehat{\phi}_{0,MS}$	111211.90 (7.32×10^6)	377.20	2.44×10^6 (1.63×10^8)	305.47	105685.89 (6.72×10^6)	427.75
100	$\widehat{\phi}_{0,LS}$	574.25 (742.44)	374.63	655.62 (1066.44)	381.43	514.99 (1005.50)	404.43
	$\widehat{\phi}_{0,LS2}$	575.22 (777.44)	368.77	604.80 (1008.68)	347.96	459.93 (841.25)	367.61
	$\widehat{\phi}_{0,Q}$	581.06 (766.41)	370.94	616.29 (1028.52)	349.43	476.54 (988.49)	377.24
	$\widehat{\phi}_{0,H}$	1581.18 (12633.49)	372.53	1949.87 (17642.45)	368.91	799.62 (8804.68)	391.39
	$\widehat{\phi}_{0,MS}$	2362.82 (60278.19)	391.28	2060.09 (19449.45)	307.12	1306.94 (39875.12)	371.96
200	$\widehat{\phi}_{0,LS}$	461.37 (321.45)	375.74	473.57 (353.23)	376.64	412.96 (107.04)	384.59
	$\widehat{\phi}_{0,LS2}$	459.51 (324.31)	372.80	453.12 (340.20)	359.66	393.11 (96.59)	367.16
	$\widehat{\phi}_{0,Q}$	461.01 (327.94)	372.39	456.51 (347.96)	358.98	396.23 (101.71)	369.52
	$\widehat{\phi}_{0,H}$	668.76 (1272.59)	374.85	662.24 (1482.46)	365.10	437.23 (211.91)	375.97
	$\widehat{\phi}_{0,MS}$	684.39 (2425.51)	390.91	679.40 (1786.14)	313.86	450.11 (412.04)	349.06
500	$\widehat{\phi}_{0,LS}$	401.91 (157.04)	371.44	409.65 (162.44)	376.79	383.51 (44.89)	373.98
	$\widehat{\phi}_{0,LS2}$	400.86 (157.27)	370.26	402.25 (159.78)	369.89	376.41 (42.95)	367.22
	$\widehat{\phi}_{0,Q}$	401.08 (157.66)	370.05	402.60 (160.69)	370.30	376.81 (43.97)	368.02
	$\widehat{\phi}_{0,H}$	450.62 (306.38)	371.50	453.77 (312.17)	368.37	388.95 (74.90)	370.33
	$\widehat{\phi}_{0,MS}$	463.89 (307.86)	388.25	418.98 (360.96)	320.31	369.89 (103.52)	339.79
1000	$\widehat{\phi}_{0,LS}$	385.97 (102.16)	369.79	388.62 (104.85)	372.47	376.72 (28.28)	372.63
	$\widehat{\phi}_{0,LS2}$	385.41 (102.20)	369.20	385.06 (103.97)	369.04	373.26 (27.64)	369.24
	$\widehat{\phi}_{0,Q}$	385.59 (102.40)	369.19	385.09 (104.34)	368.64	373.36 (28.16)	369.51
	$\widehat{\phi}_{0,H}$	407.61 (177.62)	369.72	410.37 (180.45)	371.32	379.16 (45.83)	370.77
	$\widehat{\phi}_{0,MS}$	425.02 (182.88)	388.76	367.37 (189.40)	322.91	352.90 (55.90)	339.36

1. Generate an in-control Phase I sample of size m from $P(\phi_0)$ and use this sample to estimate the process distribution, $\hat{P} = P(\hat{\phi}_0)$, where $\hat{\phi}_0$ is one of the estimators shown in Section 3.2.2 and use this estimation to compute $K(\hat{P}, \hat{\phi}_0)$ defined in Appendix.
2. Generate B bootstrap samples from the estimated in-control distribution $\hat{P}$ and compute the corresponding estimated distribution $\hat{P}^*_b = P(\hat{\phi}^*_{0,b})$ and the quantities $K(\hat{P}^*_b, \hat{\phi}^*_{0,b})$ and $K(\hat{P}, \hat{\phi}^*_{0,b})$ for $b = 1, 2, \ldots, B$.
3. Consider the bootstrap distribution of $K(\hat{P}^*_b, \hat{\phi}^*_{0,b}) - K(\hat{P}, \hat{\phi}^*_{0,b})$ and let K_{α^*} be the α^*-th quantile of that distribution. Then, the quantity $K(\hat{P}, \hat{\phi}_0) - K_{\alpha^*}$ is taken as the K-value to adjust the chart control limits.

This procedure is detailed in Algorithm 2. In Appendix it is shown that

$$K(\hat{P}, \hat{\phi}_0) = K(\hat{P}^*_b, \hat{\phi}^*_{0,b}) = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

and

$$K(\hat{P}, \hat{\phi}^*_{0,b}) = \frac{\hat{C}^*_{2,b}}{\hat{C}_2}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Therefore, finding $K(\hat{P}, \hat{\phi}_0) - K_{\alpha^*}$ is the same as finding the α^*-th quantile of the bootstrap distribution of $K(\hat{P}, \hat{\phi}^*_{0,b})$. Further, since the estimated distribution and parameters obtained in Step 1 could differ from sample to sample the bootstrapping methodology was applied several times. It has to be noted that equation (3.3.1) is valid as long as $|\phi_0| < 1$ and $|\hat{\phi}_0| < 1$. While the former is always met since the election of ϕ_0 values on the simulation study, the latter one depends on the estimation of ϕ_0 and then on the observed series. When a series does not met this constraint it is not considered and another one is generated. This procedure is stated in Algorithms 1 and 2 with the **if** condition.

Algorithm 2 Calculation of the adjusted K-values

Define m, n, ϕ_0, ARL_0, $\alpha^*, B, rep, \phi_{EST}$.
$\delta \leftarrow 0$
$\alpha \leftarrow \frac{1}{ARL_0}$
$r \leftarrow 1$
while $r \leq rep$ **do**
  Generate $X_1, X_2, \ldots, X_m$ based on the AR(1) model:
$$X_j - \mu_0 = \phi_0(X_{j-1} - \mu_0) + \varepsilon_j, \text{ where } \varepsilon_j \sim N\left(\mu_0, \sigma_0\sqrt{1-\phi_0^2}\right)$$
  Estimate ϕ_0 with ϕ_{EST} and denote it by $\hat{\phi}_{0,r}$
  if $\left|\hat{\phi}_{0,r}\right| < 1$ **then**
    $\hat{C}_2 \leftarrow$ calculate using equation (3.2.16) with $\hat{\phi}_{0,r}$.
    $b \leftarrow 1$
    while $b \leq B$ **do**
      Generate $X_1^*, X_2^*, \ldots, X_m^*$ based on the estimated $AR(1)$ model:
$$X_j - \mu_0 = \hat{\phi}_{0,r}\left(X_{j-1} - \mu_0\right) + \varepsilon_j, \text{ where } \varepsilon_j \sim N\left(\mu_0, \sigma_0\sqrt{1-\phi_0^2}\right).$$
      Estimate $\hat{\phi}_{0,r}$ with ϕ_{EST} and denote it by $\hat{\phi}_{0,b}^*$.
      if $\left|\hat{\phi}_{0,b}^*\right| < 1$ **then**
        $\hat{C}_{2,b}^* \leftarrow$ calculate using equation (3.2.16) with $\hat{\phi}_{0,b}^*$.
        $K\left(\hat{P}, \hat{\phi}_{0,b}^*\right) \leftarrow \left(\frac{\hat{C}_{2,b}^*}{\hat{C}_2}\right)\Phi^{-1}\left(1 - \frac{1}{2ARL_0}\right)$.
        $b \leftarrow b + 1$.
      end if
    end while
    $K_r \leftarrow$ is the α^*-th quantile of $K\left(\hat{P}, \hat{\phi}_{0,b}^*\right)$
    $r \leftarrow r + 1$.
  end if
end while
$K \leftarrow \text{mean}\left(K_1, \ldots, K_{rep}\right)$

First, the adjusted K-values were obtained using Algorithm 2 and then Algorithm 1 was used considering these K-values to compute the corresponding ARL's.

Table 3.4.1 shows the averaged adjusted K-values and the $ARL_{0,0.1}$, the 10-th quantile of the ARL_0's distribution with adjusted control limits. It can be seen that in all cases control limits are widened which is known to increase the ARL_0 but at the same time affects the out-of-control ARL, ARL_1. In order to evaluate the effect of the control limits adjustment, the ARL_1 for the $\bar{X}$ control chart with both unadjusted (U) and adjusted (A) control limits was obtained by using Algorithm 1 based on the corresponding adjusted K-values in Table 3.4.1. As the overestimation of ϕ_0 could lead to high ARL_1 values (see Figure 3.3.2), the 90-th quantile of the conditional distribution of the ARL_1, $ARL_{1,0.9}$, is reported in Table 3.4.2 to show that 90% of the charts will have an ARL_1 of *at most* that value. If all parameters are known, the ARL_1 is known, and it is reported at the bottom of Table 3.4.2 for $\delta = 1$. As expected, as the Phase I sample size increases the ARL_1 values decrease despite of the value of ϕ_0. However, for negative values of ϕ_0 the effect on the increase of ARL_1 values due to the control limits adjustment is less intense than for positive values of ϕ_0. Nevertheless, this effect is mitigated as the Phase I sample size increases (see Table 3.4.2). In all cases, the robust estimators have a larger $ARL_{1,0.90}$ than the least-squares based estimators.

Table 3.4.1: The averaged adjusted K values for the control limits for different values of m and ϕ_0 for a desired $ARL_0 = 370.4$ and $n = 5$. $ARL_{0,0.1}$ is the 10-th quantile of the simulated ARL_0 distribution obtained using these values.

m	$\hat{\phi}_0$	$\phi_0 = 0.9$		$\phi_0 = 0.5$		$\phi_0 = 0.1$		$\phi_0 = -0.1$		$\phi_0 = -0.5$		$\phi_0 = -0.9$	
		K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.10}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$
	$\widehat{\phi}_{0,LS}$	3.37	325.39	3.47	352.22	3.48	369.19	3.47	370.95	3.40	363.06	3.13	493.21
	$\widehat{\phi}_{0,LS2}$	3.28	411.91	3.44	388.47	3.49	359.30	3.50	338.45	3.47	287.33	3.12	440.40
50	$\widehat{\phi}_{0,Q}$	3.31	377.80	3.46	369.21	3.50	351.41	3.51	326.92	3.48	285.72	3.14	449.76
	$\widehat{\phi}_{0,H}$	3.48	231.82	3.70	190.22	3.78	177.94	3.79	165.46	3.67	156.12	3.20	440.36
	$\widehat{\phi}_{0,MS}$	3.53	193.49	3.72	178.96	3.83	161.65	3.89	127.86	3.77	107.58	3.23	432.96
	$\widehat{\phi}_{0,LS}$	3.22	354.24	3.32	374.76	3.33	370.62	3.32	370.73	3.29	362.74	3.07	409.29
	$\widehat{\phi}_{0,LS2}$	3.17	403.16	3.29	398.00	3.33	368.88	3.34	357.27	3.33	324.13	3.08	384.98
100	$\widehat{\phi}_{0,Q}$	3.18	384.45	3.30	387.22	3.33	361.97	3.34	353.86	3.33	319.13	3.08	383.26
	$\widehat{\phi}_{0,H}$	3.29	286.92	3.47	239.71	3.53	213.00	3.54	202.06	3.49	200.77	3.11	371.58
	$\widehat{\phi}_{0,MS}$	3.31	245.93	3.49	217.23	3.55	217.61	3.60	169.42	3.61	140.66	3.13	363.67
	$\widehat{\phi}_{0,LS}$	3.13	369.43	3.21	367.47	3.23	368.43	3.23	370.07	3.21	371.12	3.05	383.11
	$\widehat{\phi}_{0,LS2}$	3.11	395.34	3.20	379.75	3.23	368.66	3.23	364.40	3.23	351.63	3.05	370.31
200	$\widehat{\phi}_{0,Q}$	3.12	386.98	3.20	376.00	3.23	367.96	3.23	361.67	3.23	349.76	3.05	368.50
	$\widehat{\phi}_{0,H}$	3.18	313.07	3.32	266.58	3.36	253.98	3.37	246.06	3.35	242.36	3.08	353.35
	$\widehat{\phi}_{0,MS}$	3.19	295.03	3.33	258.98	3.37	261.72	3.40	227.43	3.46	177.78	3.09	336.76
	$\widehat{\phi}_{0,LS}$	3.07	370.35	3.13	372.15	3.14	372.80	3.14	367.72	3.14	365.56	3.03	372.38
	$\widehat{\phi}_{0,LS2}$	3.06	380.77	3.13	377.50	3.14	373.31	3.14	365.85	3.14	358.11	3.03	366.80
500	$\widehat{\phi}_{0,Q}$	3.07	378.35	3.13	376.21	3.14	372.97	3.14	365.95	3.14	358.32	3.03	365.97
	$\widehat{\phi}_{0,H}$	3.11	333.92	3.20	297.55	3.22	286.81	3.23	284.91	3.22	283.42	3.05	351.15
	$\widehat{\phi}_{0,MS}$	3.09	336.8	3.19	304.44	3.23	280.97	3.23	280.24	3.31	224.28	3.06	322.63
	$\widehat{\phi}_{0,LS}$	3.05	371.86	3.09	369.34	3.10	369.80	3.10	370.91	3.10	370.29	3.02	370.80
	$\widehat{\phi}_{0,LS2}$	3.04	377.12	3.09	372.07	3.09	370.14	3.10	370.07	3.10	366.60	3.02	367.85
1000	$\widehat{\phi}_{0,Q}$	3.05	376.16	3.09	371.64	3.10	369.37	3.10	370.20	3.10	366.24	3.03	367.17
	$\widehat{\phi}_{0,H}$	3.07	344.85	3.14	319.37	3.16	307.77	3.16	310.49	3.16	306.34	3.04	354.40
	$\widehat{\phi}_{0,MS}$	3.05	364.39	3.12	329.66	3.16	290.75	3.16	308.68	3.23	246.59	3.05	321.86

3.5 Conclusions and future work

As the autocorrelation degrades the performance of control charts, the relationship between the conditional in-control and out-of-control ARLs and the estimation of the autoregressive parameter ϕ_0 was presented. Results show that overestimation of ϕ_0 provides higher ARL values for either the process being under control or not. When $\phi_0 < 0$ it is more likely to have larger ARL_0 values and small ARL_1 ones. However, even for $\phi_0 < 0$ around 50% of the times control charts have a smaller ARL_0 than the expected one, and it was shown that control limits adjustment to guarantee an in-control performance had little impact on ARL_1.

As the overestimation of ϕ_0 affects the ARL of the $\bar{X}$ chart, it is recommended to use an estimator with the smaller bias and standard deviation. It does not seem that one estimator dominates the other ones in all cases, but $\hat{\phi}_{0,LS}$ performs better in most of them. Even though $\hat{\phi}_{0,LS}$ is outperformed in some scenarios the difference between estimators becomes small as m increases.

The evaluation of the $\bar{X}$ control chart performance in presence of outliers where the robust estimators might have better performance is left as future work as well as the extension to AR processes of higher orders, Moving Averages (MA) and ARMA processes. Also the evaluation of the EWMA and CUSUM control charts. On the other hand, recently Zwetsloot and Woodall (2017) provided a definition to standardize the meaning of *better conditional performance* among two charts, when they are compared head-to-head. This definition alongside with the $AARL$ and $SDARL$ metrics might be considered to reevaluate the charts performance.

Appendix

Here it is explained how to calculate the adjusted K-values in order to obtain the adjusted control limits for the Shewhart $\bar{X}$ control chart for AR(1) processes. The statistic under control is the sample mean from

Table 3.4.2: $ARL_{1,0.9}$ is the 90-th quantile of the simulated ARL_1 values considering a shift of $\delta = 1$, with $n = 5$,$ARL_0 = 370.4$

m	$\hat{\phi}$	$\phi_0 = 0.9$		$\phi_0 = 0.5$		$\phi_0 = 0.1$		$\phi_0 = -0.1$		$\phi_0 = -0.5$		$\phi_0 = -0.9$	
		U	A	U	A	U	A	U	A	U	A	U	A
50	$\widehat{\phi}_{0,LS}$	42.34	111.50	30.38	112.80	12.11	37.55	6.95	18.42	2.19	3.53	1.07	1.10
	$\widehat{\phi}_{0,LS2}$	52.91	148.61	35.53	140.61	12.82	40.76	7.05	18.84	2.11	3.35	1.06	1.08
	$\widehat{\phi}_{0,Q}$	49.69	138.06	36.94	144.70	13.11	42.38	7.19	19.39	2.16	3.42	1.07	1.09
	$\widehat{\phi}_{0,H}$	52.80	145.63	57.17	271.41	21.65	87.70	10.93	40.02	2.86	5.32	1.10	1.13
	$\widehat{\phi}_{0,MS}$	56.37	160.10	87.01	478.17	26.45	132.92	12.23	41.99	3.00	5.64	1.12	1.16
100	$\widehat{\phi}_{0,LS}$	43.17	74.83	24.73	55.59	9.49	18.75	5.47	9.87	1.90	2.51	1.05	1.06
	$\widehat{\phi}_{0,LS2}$	48.24	85.12	26.47	60.66	9.68	19.27	5.49	9.91	1.87	2.45	1.05	1.05
	$\widehat{\phi}_{0,Q}$	48.68	86.06	26.72	62.36	9.82	19.59	5.54	9.92	1.88	2.47	1.05	1.06
	$\widehat{\phi}_{0,H}$	52.74	92.80	37.99	93.34	13.75	29.49	7.55	14.96	2.18	3.13	1.06	1.07
	$\widehat{\phi}_{0,MS}$	56.98	103.68	49.80	131.88	15.43	34.00	7.39	14.70	2.23	3.30	1.06	1.08
200	$\widehat{\phi}_{0,LS}$	42.26	59.01	21.48	35.61	8.05	12.24	4.79	6.80	1.74	2.07	1.04	1.05
	$\widehat{\phi}_{0,LS2}$	44.62	62.65	22.15	36.93	8.11	12.36	4.79	6.79	1.73	2.05	1.04	1.04
	$\widehat{\phi}_{0,Q}$	45.20	63.31	22.28	37.12	8.13	12.37	4.80	6.83	1.73	2.06	1.04	1.04
	$\widehat{\phi}_{0,H}$	48.79	69.04	27.78	47.97	10.08	16.14	5.82	8.63	1.91	2.33	1.05	1.05
	$\widehat{\phi}_{0,MS}$	56.82	81.65	34.44	63.27	10.66	17.00	5.70	8.29	1.91	2.34	1.05	1.05
500	$\widehat{\phi}_{0,LS}$	40.44	48.71	18.91	24.97	7.07	9.00	4.25	5.17	1.64	1.81	1.04	1.04
	$\widehat{\phi}_{0,LS2}$	41.31	49.82	19.12	25.28	7.09	9.03	4.25	5.17	1.63	1.80	1.04	1.04
	$\widehat{\phi}_{0,Q}$	41.51	50.07	19.16	25.30	7.10	9.03	4.25	5.16	1.64	1.80	1.04	1.04
	$\widehat{\phi}_{0,H}$	44.26	53.49	21.67	29.76	8.07	10.59	4.74	5.90	1.73	1.92	1.04	1.04
	$\widehat{\phi}_{0,MS}$	54.13	66.16	25.76	34.92	8.18	10.69	4.79	5.97	1.71	1.90	1.04	1.04
1000	$\widehat{\phi}_{0,LS}$	39.48	44.34	17.63	21.38	6.61	7.78	4.02	4.61	1.59	1.70	1.03	1.04
	$\widehat{\phi}_{0,LS2}$	39.90	44.83	17.73	21.50	6.62	7.79	4.02	4.61	1.59	1.70	1.03	1.04
	$\widehat{\phi}_{0,Q}$	39.97	44.91	17.75	21.45	6.62	7.79	4.02	4.61	1.59	1.70	1.03	1.04
	$\widehat{\phi}_{0,H}$	41.89	47.22	19.61	23.66	7.22	8.58	4.34	5.00	1.65	1.77	1.04	1.04
	$\widehat{\phi}_{0,MS}$	50.72	57.55	22.19	27.01	7.19	8.58	4.44	5.14	1.63	1.75	1.04	1.04
Known case values		36.12		15.00		5.72		3.54		1.50		1.03	

subgroups of size n drawn from the process. Assuming an in-control process ($\delta = 0$), where μ_0 and σ_0 are known parameters, we calculate $K(P, \hat{\phi}_0)$ as the solution of $1 - \alpha = \mathbb{P}(\widehat{LCL} < \bar{X}_i < \widehat{UCL} | \hat{\phi}_0)$, i.e. , the value that solves:

$$1 - \alpha = \Phi\left(K\frac{C_2}{\hat{C}_2}\right) - \Phi\left(-K\frac{C_2}{\hat{C}_2}\right) \tag{3.5.1}$$

where α is the probability of the chart to signal an out-of-control observation. Using the notation introduced in Section 3.4 where $P = P(\phi_0)$ refers to the autoregressive model in equation (3.2.8) with $\delta = 0$ and autoregressive parameter ϕ_0, then $K(P, \hat{\phi}_0)$ depends upon P and $\hat{\phi}_0$ in the following manner: C_2 in equation (3.5.1) is calculated using the autoregressive parameter from the model P whereas $\hat{C}_2$ by using $\hat{\phi}_0$. Note that $\hat{P} = P(\hat{\phi}_0)$ and $\hat{P}_b^* = P(\hat{\phi}_{0,b}^*)$. Therefore, we can calculate $K(\hat{P}, \hat{\phi}_0)$, $K(\hat{P}_b^*, \hat{\phi}_{0,b}^*)$ and $K(\hat{P}, \hat{\phi}_{0,b}^*)$. Considering first $K(\hat{P}, \hat{\phi}_0)$, as the autoregressive parameter from model $\hat{P}$ is $\hat{\phi}_0$ then $K(\hat{P}, \hat{\phi}_0)$ is the value of K such that $1 - \alpha = \Phi(K) - \Phi(-K) = 2\Phi(K) - 1$, i.e.

$$K = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

and that is also true for $K(\hat{P}_b^*, \hat{\phi}_{0,b}^*)$ since the autoregressive parameter in model $\hat{P}_b^*$ is $\hat{\phi}_{0,b}^*$. Finally, in order to obtain $K(\hat{P}, \hat{\phi}_{0,b}^*)$, we have to find the K-value satisfying:

$$1 - \alpha = \Phi\left(K\frac{\hat{C}_2}{\hat{C}_{2,b}^*}\right) - \Phi\left(-K\frac{\hat{C}_2}{\hat{C}_{2,b}^*}\right) = 2\Phi\left(K\frac{\hat{C}_2}{\hat{C}_{2,b}^*}\right) - 1,$$

i.e.

$$K = \left(\frac{\hat{C}_{2,b}^*}{\hat{C}_2}\right)\Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Chapter 4. The conditional performance of the $\bar{X}$ control chart for AR(1) processes under estimated parameters

Abstract

Control charts are Statistical Process Monitoring tools known to be helpful for the detection of assignable causes of variation. Despite their ability to identify departures from in-control situations, their sensitivity can be negatively affected by the violation of the independence assumption of observations and the lack of knowledge of the true values of the process parameters. Control charts are usually evaluated by means of the Average Run Length, ARL, which is a random variable when process parameters are estimated. In that case, the Average and Standard Deviation of the ARL, ($AARL$ and $SDARL$, respectively), are the suggested performance measures since they capture the sampling variability of the ARL due to Phase I estimations. In this research, the performance of the $\bar{X}$ control chart for monitoring the mean of stationary AR(1) processes is evaluated by considering different estimators for the process parameters. A bootstrap-based method is applied to adjust the $\bar{X}$ control limits in order to have a guaranteed in-control performance and its impact on the out-of-control ARL is also presented. A numerical example based on a real dataset is provided in order to illustrate how the chart and the techniques developed here are implemented in practice.

Keywords: AR(1) process; $\bar{X}$ control chart; estimated parameters; guaranteed performance.

4.1 Introduction

Control charts are Statistical Process Monitoring (SPM, also called Statistical Process Control or SPC) tools for monitoring and detecting changes in process parameters. Their performance depends upon the validation of assumptions made at the design stage, such as the distribution of the data, independence of observations, or prior knowledge of the process in-control parameters, among others. When these assumptions are not met, control charts might no longer be suitable to be implemented in practice. Their performance might have an unexpected behavior, unless a modification is made to their design.

There are many practical situations where assuming independence might not be appropriate due to a significant level of autocorrelation in the process. These situations are commonly found, for example, in health surveillance and crop monitoring, where measurements are taken from the same object; chemical processes, where a continuous flow is monitored; or processes where observations are collected within small periods of time. It is known that the implementation of the traditional control charts, built for independent observations, is not advisable, as serial correlation has a significant effect on performance (Psarakis et al., 2014). Nevertheless, there are some techniques to improve the charts performance when dealing with these situations, or even to develop new control charts, as Psarakis and Papaleonida (2007) and Prajapati and Singh (2012) pointed out in their literature reviews.

A widely known approach to handle autocorrelated data is the time-series approach by Alwan and Roberts (1988), where a time-series model is fitted to the observed data, and the corresponding residuals are assessed with a control chart scheme that assumes independence. Trying to improve the performance of these charts for residuals, Triantafyllopoulos and Bersimis (2016) proposed a modification based on a Bayes factor to detect departures from in-control situations, making a comparison with the former. Recently, Dawod et al. (2017) evaluated the performance of the Shewhart, EWMA and CUSUM charts for residuals over AR(1), MA(1) and ARMA(1,1) processes by considering well-fitted models with known parameters.

Different developments were carried by Lwin (2011) who considered the estimation of the autoregressive parameter to set up the EWMAST control chart for AR(1) processes with measurement errors in the variables. Alshraideh and Khatatbeh (2014) used the multivariate normal distribution to introduce the Gaussian Process control chart, and Zhang and Pintar (2015) considered the Exponentially Weighted Mean Square (EWMS) control chart for monitoring the variance of stationary processes. Wu (2016) considered the change-point problem for AR(1) processes monitored with a model-based CUSUM control chart. Dasdemir et al. (2016) evaluated the effect of outliers on the Phase II performance of a modified AR(1) Shewhart chart. Osei-Aning et al. (2017b) provided an optimal scheme for monitoring the mean of stationary AR(1) processes for the CUSUM and EWMA control charts by finding the parameters that minimize the extra quadratic loss function over a range of possible changes. Finally, Osei-Aning et al. (2017a) introduced the mixed EWMA-CUSUM and CUSUM-EWMA control charts for autocorrelated processes.

In addition to the fact that the observations might not be considered independent, in-control process parameters are often *unknown*. A lot of research work has been done over several scenarios, including independent data, autocorrelated data and attribute data, among others. Most of these works have been devoted to compare estimators for the mean and/or the variance, and to evaluate the charts performance using measures related to the Run Length RL, such as the Average Run Length (ARL), its standard deviation ($SDRL$), or some quantiles of the RL's distribution. Readers interested in this topic are referred to the literature reviews from Jensen et al. (2006) and Psarakis et al. (2014).

Recently, two topics about the evaluation and design of control charts under estimated parameters have attracted the attention of SPM researchers: the first one is the 'practitioner-to-practitioner' variability of the ARL due to Phase I estimations: Jones and Steiner (2012) introduced the Standard Deviation of the ARL ($SDARL$) as a performance measure to account for this variability. The second one is the guaranteed in-control performance: Gandy and Kvaløy (2013) proposed a bootstrapping method to adjust the control chart limits to have a guaranteed performance with a certain probability. There are other techniques to guarantee an in-control performance, such as the adjustment proposed by Albers and Kallenberg (2005), the use of tolerance limits or exacts methods developed for specific charts (e.g. Goedhart et al. 2017a, Goedhart et al. 2017b, and Faraz et al. 2018). Following these results, the evaluation and design of several control charts have been revisited.

Within control charts revisions, for attribute data, we have Zhang et al. (2013), Lee et al. (2013), Zhao and Driscoll (2016) and Faraz et al. (2017); control charts for monitoring the mean by Zhang et al. (2014), Saleh et al. (2015a), Saleh et al. (2015b), Aly et al. (2015b), Aly et al. (2016), Jeske (2016b), Saleh et al. (2016), Hany and Mahmoud (2016), Hu and Castagliola (2017) and Goedhart et al. (2017b); control charts for monitoring the variance by Epprecht et al. (2015), Faraz et al. (2015), Diko et al. (2017), Goedhart et al. (2017a), Guo and Wang (2017); the self-starting methodology by Keefe et al. (2015); the linear profiles by Aly et al. (2015a) and monitoring of time between events by Cheng et al. (2018). Moreover, the Shewhart, EWMA and CUSUM control charts were compared by Zwetsloot and Woodall (2017) using a definition for a 'better performance' introduced by the authors: if the ARL's values of each chart lies within 5% of the nominal ARL value they are considered as to have an equivalent performance; otherwise, the one with the ARL closer to the nominal ARL value is considered to have better performance. With this definition, they concluded that the EWMA and CUSUM charts do not have an equivalent performance under parameter estimation in opposition to the findings of Hawkins and Wu (2014) for the known parameter case.

Considering the literature reviews of Psarakis and Papaleonida (2007) and Prajapati and Singh (2012), around 251 works were reviewed under the label of monitoring autocorrelated processes whereas the literature reviews for the effect of parameter estimation due to Jensen et al. (2006) and Psarakis et al. (2014) considered around 159 works. However, only 14 of them considered the monitoring of autocorrelated processes. Then, the effect of parameter estimation on charts performance for this kind of processes has not been thoroughly studied as its counterpart for independent observations. Runger and Willemain (1995), Snoussi et al. (2005), and Dawod et al. (2017) evaluated the performance of *residuals* control charts assuming that 'perfectly accurate' parameters have been estimated; and apart from *residuals* charts, Lu and Reynolds Jr. (1999) and Lu and Reynolds Jr. (2001) provided limited simulation studies about the effect of parameter estimation. Now, considering performance measures related to the ARL distribution and/or the guaranteed conditional in-control performance topics, 25 works were found but none of them was devoted to autocorrelated processes.

With this in mind, in this study, the performance of the $\bar{X}$ control chart for monitoring the mean of stationary AR(1) processes under estimated parameters is evaluated by considering several estimators for the process mean, variance and the autoregressive parameter, and using $AARL$ and $SDARL$ as performance measures. In addition, the bootstrapping methodology of Gandy and Kvaløy (2013) is applied to adjust the control limits to ensure, with a certain probability, a conditional in-control performance. The impact of this adjustment on the out-of-control ARL (ARL_1) is also discussed.

The remaining of this chapter is organized as follows: Section 4.2 introduces the design of the $\bar{X}$ control chart for AR(1) processes dealing with *unknown* parameters, and several estimators used in the design process are introduced. In Section 4.2 the marginal relationship of the ARL with estimated parameters is shown along the approach and algorithm used to compute the $AARL$ and $SDARL$. At the end of the section, numerical results in the form of $AARL$ and $SDARL$ over several in-control scenarios are presented. Section 4.4 is devoted to the adjustment of the control limits and the impact on the ARL_1, while a numerical example is considered in Section 4.5. Conclusions and future work are discussed in Section 4.6.

4.2 Design of the Shewhart $\bar{X}$ chart for AR(1) processes under estimated parameters

The study of the $\bar{X}$ control chart for AR(1) processes was first considered by Costa and Castagliola (2011) whereas the effect of parameter estimation was addressed in Chapter 3 *but* assuming that only ϕ_0 was estimated. Now, the process mean μ_0 and process variance σ_0^2 are considered as *unknown* and therefore several estimators are considered for the study.

The same model stated in Chapter 3 is considered through this chapter, but it is stated here again for the sake of completeness. Let us assume that the observations $\{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\}$ follow the AR(1) model, given by

$$X_{i,j} - \mu_0 - \delta\sigma_0 = \phi_0 \left(X_{i,j-1} - \mu_0 - \delta\sigma_0\right) + \varepsilon_{i,j}, \tag{4.2.1}$$

for $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, n$; where $\varepsilon_{i,j} \sim N\left(0, \sigma_\varepsilon\right)$ is a normal random noise; μ_0, σ_0 and $\phi_0 \in (-1, 1)$ are the in-control process mean, process variance and process autoregressive parameter, respectively. Further, $\delta = \frac{|\mu_0 - \mu_1|}{\sigma_0}$ is the standardized mean shift from μ_0 to μ_1 measured in standard deviations. The process is said to be under statistical control whenever $\delta = 0$; otherwise, it is considered to be out-of-control. Furthermore, under model 4.2.1 the process variance is related to the errors variance by means of

$$\sigma_0^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_0^2}. \tag{4.2.2}$$

The sample statistic used by the Shewhart $\bar{X}$ control chart for monitoring the mean μ_0 of AR(1) processes is the sample mean $\bar{X}_i = \frac{1}{n}(X_{i,1} + X_{i,2} + \cdots + X_{i,n})$. A widely used sampling strategy calls for collecting consecutive parts or cycles as rational subgroups (Wheeler, 2015), which can lead to autocorrelation within samples. If the sampling interval is sufficiently large, then autocorrelation between consecutive samples is neglegible, and, for $j = 1, 2, \ldots, n$, $X_{i,j}$ and $X_{i+1,j}$ can be considered independent. It is noteworthy to say that, under model stated in equation 4.2.1, the variable $X_{i,0}$ is non-observable, and, as the sampling interval is large, it can be assumed to follow the steady-state distribution $N\left(\mu_0, \sigma_0\right)$. In this case, Alwan and Radson (1992) found that the standard deviation of the sample mean $\sigma(\bar{X}_i)$ of the marginal distribution of stationary AR(1) processes ($|\phi_0| < 1$) is given by

$$\sigma(\bar{X}_i) = \frac{\sigma_0}{\sqrt{n}C_2}, \tag{4.2.3}$$

where C_2 is a function of n and ϕ_0 defined as

$$C_2 = C_2\left(n, \phi_0\right) = \sqrt{\frac{n}{n + 2\left(\frac{\phi_0^{n+1} - n\phi_0^2 + (n-1)\phi_0}{(\phi_0 - 1)^2}\right)}}. \tag{4.2.4}$$

Note that equation 4.2.4 is valid whenever $|\phi_0| < 1$.

When the process parameters are unknown, a retrospective Phase I implementation of the control chart is performed to get reliable estimates, and, at the same time, to assess process stability. In this research, as we are assuming that the autocorrelation is within but not between samples, the Phase I is done by collecting m (larger than n) consecutive observations over a sufficiently large period of time to check for process stability and also to capture all the information in the sample about the process parameters. After that, the $\bar{X}$ control chart is set up for Phase II, where each sample consists of n consecutive autocorrelated observations. Samples are collected over sufficiently long time intervals (for example, every h hours) in order to reduce the autocorrelation between samples to the point where it is suitable to assume that the sample means are independent. Under this approach, consider that, within the Phase I sample, observations follow the model

$$X_j - \mu_0 = \phi_0 \left(X_{j-1} - \mu_0\right) + \varepsilon_j, \tag{4.2.5}$$

for $j = 1, 2, \ldots, m$, where $X_0 \sim N\left(\mu_0, \sigma_0\right), \varepsilon_j \sim N\left(0, \sigma_\varepsilon\right)$ and $|\phi_0| < 1$.

When implementing the Shewhart $\bar{X}$ control chart for autocorrelated processes with estimated parameters, the control limits become

$$\widehat{LCL} = \hat{\mu}_0 - K \frac{\hat{\sigma}_0}{\sqrt{n}\hat{C}_2}, \tag{4.2.6}$$

$$\widehat{UCL} = \hat{\mu}_0 + K \frac{\hat{\sigma}_0}{\sqrt{n}\hat{C}_2}, \tag{4.2.7}$$

with

$$\hat{C}_2 = C_2\left(n, \hat{\phi}_0\right) = \sqrt{\frac{n}{n + 2\left(\frac{\hat{\phi}_0^{n+1} - n\hat{\phi}_0^2 + (n-1)\hat{\phi}_0}{(\hat{\phi}_0 - 1)^2}\right)}}, \tag{4.2.8}$$

where $\hat{\mu}_0$, $\hat{\sigma}_0$ and $\hat{\phi}_0$ are estimators (to be defined) of μ_0, σ_0 and ϕ_0, respectively.

In order to set up the $\bar{X}$ control chart with estimated parameters, let us consider a Phase I sample of consecutive observations $X_1, X_2, \ldots, X_m$, where the estimator of the *process mean* μ_0 is the sample mean

$$\hat{\mu}_0 = \bar{X} = \frac{1}{m}\sum_{j=1}^{m} X_j. \tag{4.2.9}$$

Concerning the *process standard deviation* σ_0, four different estimators are considered here. The first one is the sample standard deviation

$$\hat{\sigma}_{0,SQ1} = S = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}\left(X_j - \bar{X}\right)^2}, \tag{4.2.10}$$

which is commonly used in the i.i.d. case. Under the normal i.i.d. assumption, an unbiased estimator of the standard deviation is

$$\hat{\sigma}_{0,UN} = \frac{S}{c_4\left(m\right)}, \tag{4.2.11}$$

where $c_4\left(m\right) = \sqrt{\frac{2}{m-1}}\frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)}$. Another estimator based on the sums of the squared deviations from the mean is the one used in time series analysis

$$\hat{\sigma}_{0,SQ} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(X_j - \bar{X}\right)^2}. \tag{4.2.12}$$

In addition, a fourth estimator for the process standard deviation is the moving range, sometimes used in the i.i.d. case

$$\hat{\sigma}_{0,MR} = \frac{\overline{MR}}{1.128} = \frac{0.8865}{m-1}\sum_{j=2}^{m}|X_j - X_{j-1}|. \tag{4.2.13}$$

Concerning the *autoregressive parameter,* ϕ_0, the five estimators introduced in Chapter 3 are again considered in this study. They are mentioned here for the sake of completeness. The least-squares estimator

$$\hat{\phi}_{0,LS} = \frac{\sum_{j=2}^{m} X_j X_{j-1}}{\sum_{j=1}^{m-1} X_j^2}, \tag{4.2.14}$$

a less biased least squares based estimator

$$\hat{\phi}_{0,LS1} = \hat{\phi}_{0,LS}\left(1 - \frac{2}{m} + \frac{4}{m^2}\right)^{-1} = \left(\frac{m^2}{m^2 - 2m + 4}\right)\hat{\phi}_{0,LS}, \tag{4.2.15}$$

the unbiased estimator given by

$$\hat{\phi}_{0,Q} = 2\hat{\phi}_{0,LS} - \frac{1}{2}\left(\frac{\sum_{j=2}^{\lfloor m/2\rfloor} X_j X_{j-1}}{\sum_{j=1}^{\lfloor m/2\rfloor - 1} X_j^2} + \frac{\sum_{j=\lfloor m/2\rfloor+2}^{m} X_j X_{j-1}}{\sum_{j=\lfloor m/2\rfloor+1}^{m-1} X_j^2}\right), \tag{4.2.16}$$

where $\lfloor\cdot\rfloor$ is the rounded down integer function is considered. In addition to those estimators, the estimator based on the median of the moving ratios

$$\hat{\phi}_{0,H} = \text{median}\left(\frac{X_2}{X_1}, \frac{X_3}{X_2}, \ldots, \frac{X_m}{X_{m-1}}\right), \tag{4.2.17}$$

and the mean substitute estimator, $\hat{\phi}_{0,MS}$, given as the solution to the quadratic equation

$$\text{sign}(\phi)\, 0.26\phi^2 + 0.195\phi - 0.4705 \times \frac{\text{median}(X_1X_2, \ldots, X_{m-1}X_m)}{\text{median}(X_1^2, X_2^2, \ldots, X_{m-1}^2)} = 0. \tag{4.2.18}$$

Once the parameters are estimated, the Phase II implementation of the $\bar{X}$ control chart can be started to monitor the mean of stationary AR(1) processes by using the control limits shown in (4.2.6) and (4.2.7). The performance of the $\bar{X}$ chart for $AR(1)$ processes with estimated parameters is evaluated in the next section.

4.3 Conditional in-control performance of the $\bar{X}$ chart for an AR(1) process under parameter estimation

The $AARL$ and $SDARL$ can be computed from the ARL distribution or by running extensive Monte Carlo simulations in order to increase the estimation accuracy. In the i.i.d. case, several authors (Saleh et al., 2015a, Saleh et al., 2015b, Saleh et al., 2016 and Zwetsloot and Woodall, 2017) have considered the first approach, i.e., the use of the ARL's distribution. They have rewritten the chart statistic and they have used the fact that $\bar{X}$ and S^2 are independent in order to find the ARL's distribution. The integrals involved in the expectations have been obtained numerically by using Gaussian quadrature methods. Even though the statistic might be rewritten in a similar way for AR(1) processes, the independence of $\bar{X}$ and S^2 does not hold any longer, (Schöne and Schmid, 2000), and it is not possible to obtain the p.d.f. of the ARL using the same approach as in Zwetsloot and Woodall (2017). As a consequence, we used Monte Carlo simulations to compute the $AARL$, $SDARL$ and other quantiles of the ARL distribution.

The conditional probability of not getting an out-of-control signal in the $\bar{X}$ control chart for AR(1) processes is given by

$$\beta = \mathbb{P}(\bar{X}_i \in [\widehat{LCL}, \widehat{UCL}] | \hat{\mu}_0, \hat{\sigma}_0, \hat{\phi}_0),$$

and, after some manipulations, this probability becomes

$$\beta = \Phi\left(\frac{\sqrt{n}C_2}{\sigma_0}\left(\hat{\mu}_0 - \mu_0 - \delta\sigma_0\right) + K\frac{\hat{\sigma}_0 C_2}{\sigma_0 \hat{C}_2}\right) - \Phi\left(\frac{\sqrt{n}C_2}{\sigma_0}\left(\hat{\mu}_0 - \mu_0 - \delta\sigma_0\right) - K\frac{\hat{\sigma}_0 C_2}{\sigma_0 \hat{C}_2}\right), \tag{4.3.1}$$

where μ_0, σ_0 and ϕ_0 denote the true in-control process parameters and $\hat{\mu}_0, \hat{\sigma}_0$ and $\hat{\phi}_0$ are their respective estimations obtained from the retrospective Phase I study. Conditioned on these estimations, sample means $\bar{X}_i$'s are assumed to be independent, and, therefore, the run length of the $\bar{X}$ control chart follows a geometric distribution with parameter $1-\beta$ and mean $ARL = 1/(1-\beta)$.

We first examine the marginal effect of each parameter on the ARL, that is, we investigate the effect of the estimation of only one parameter when all other parameters are fixed at their true in-control values, with n, K and δ also fixed at some constant value. For instance, if we want to see the marginal effect of the estimation of the process mean, we will consider that $\hat{\sigma}_0 = \sigma_0$ and $\hat{\phi}_0 = \phi_0$, (which implies that $\hat{C}_2 = C_2$), for some fixed ϕ_0 values in equation (4.3.1). For convenience, we will assume $n = 5$, $ARL_0 = 370.4$, $\delta = 1$ for the ARL_1 and $\phi_0 \in \{\pm 0.1, \pm 0.5, \pm 0.9\}$ in order to explore low, medium and high levels of both positive and negative autocorrelation. In addition, and without loss of generality, we consider normal observations with $\mu_0 = 0$ and $\sigma_0 = 1$. In this case, equation (4.3.1) reduces to

$$\beta = \Phi\left(\sqrt{5}C_2\left(\hat{\mu}_0 - \delta\right) + 3\right) - \Phi\left(\sqrt{5}C_2\left(\hat{\mu}_0 - \delta\right) - 3\right). \tag{4.3.2}$$

Figure 4.3.1 (for $\delta = 0$) and Figure 4.3.2 (for $\delta = 1$) show the relationship between the ARL and $\hat{\mu}_0$ where the ARL was calculated using equation (4.3.2). From these figures it can be inferred that (i) the ARL is symmetrical with respect to $\hat{\mu}_0 = \delta$ (in general, with respect to $\mu_0 + \delta\sigma_0$) for both in-control and out-of-control ARL no matter the true value of ϕ_0, and (ii) the ARL tends to be larger for positive autocorrelation than for negative one. The marginal effect on the ARL due to the estimation of σ_0 is shown in Figures 4.3.3 and 4.3.4 for $\delta = 0$ and $\delta = 1$, respectively. Note that Figure 4.3.3 only has one curve, since, in that particular set up, $\hat{\mu}_0 = \mu_0, \delta = 0$ and $\hat{\phi}_0 = \phi_0$ (which implies $\hat{C}_2 = C_2$) so that C_2 is not involved in equation (4.3.1) and, therefore, in the calculation of the ARL_0. It can be seen that (iii) for a fixed value of $\hat{\sigma}_0$ the ARL_1 tends to be larger for positive autocorrelation than for negative one, and (iv) overestimation of σ_0 leads to higher ARL values. Finally, recall that the marginal effect on the ARL due to ϕ_0 estimation was addressed in Chapter 3, and it is depicted in Figures 3.3.1 and 3.3.2, for $\delta = 0$ and $\delta = 1$, respectively. The plots show that (v) overestimation of ϕ_0 leads to larger ARL values.

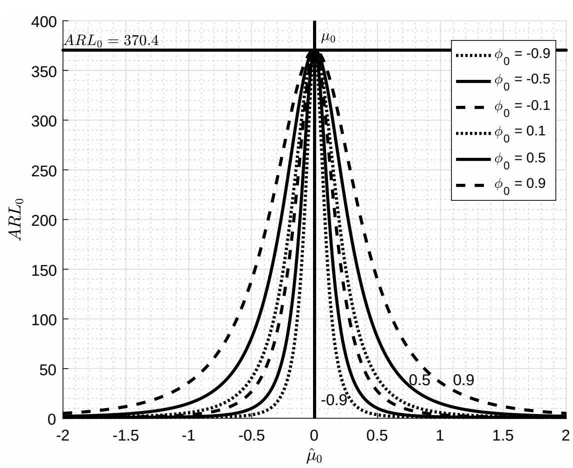


Figure 4.3.1: ARL_0 vs $\hat{\mu}_0$, for $n = 5$, $\delta = 0$,(in-control process)

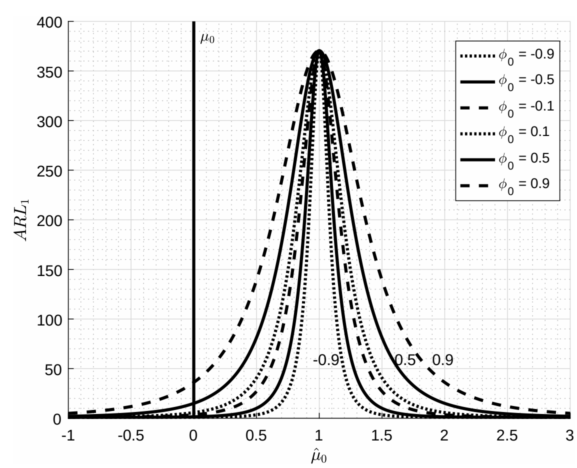


Figure 4.3.2: ARL_1 vs $\hat{\mu}_0$, for $n = 5$, $\delta = 1$,(out-of-control process)

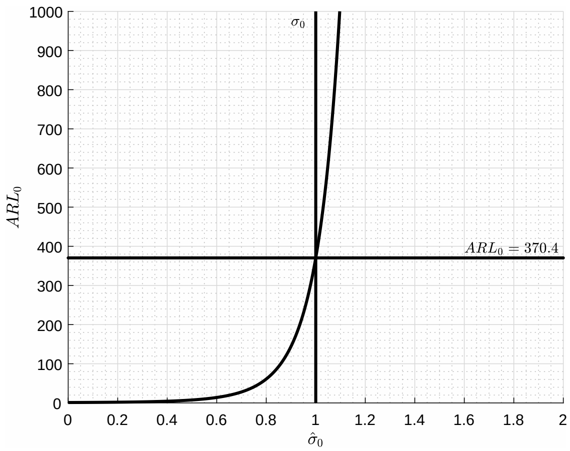


Figure 4.3.3: ARL_0 vs $\hat{\sigma}_0$, for $n = 5, \delta = 0$, (in-control process)

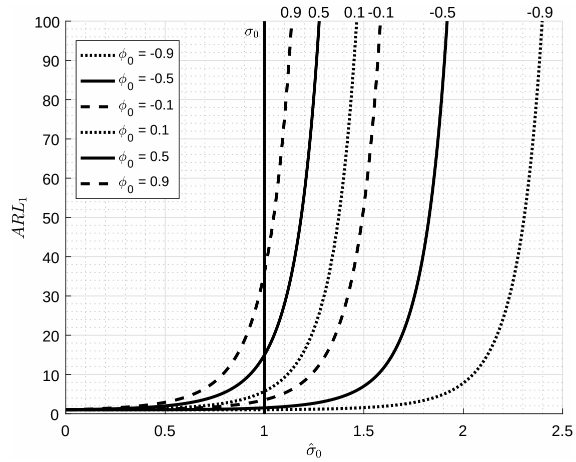


Figure 4.3.4: ARL_1 vs $\hat{\sigma}_0$, for $n = 5$, $\delta = 1$, (out-of-control process)

Results (i) to (v) are explained here by considering first that, when n is fixed, $C_2 = C_2(n, \phi_0)$ is a decreasing function of ϕ_0. In fact, $C_2(n, \phi_0) > 1$ if $\phi_0 < 0$ and $0 < C_2(n, \phi_0) < 1$ if $\phi_0 > 0$. This means that there is a widening (shrinking) of the control limits for positive (negative) autocorrelation. Now, concerning *the marginal effect of estimating* μ_0, the estimated control limits used are $\widehat{UCL}/\widehat{LCL} = \hat{\mu}_0 \pm K\frac{\sigma_0}{\sqrt{n}C_2}$, and the widening (or shrinking) of the control limits is around the line $\hat{\mu}_0$. Hence, if $\hat{\mu}_0 \neq \mu_0$, these control limits will move away from the true center line μ_0, but the proportion of observations lying within the control limits will be (i) the same, symmetrically around $\mu_0 + \delta\sigma_0$ since $\bar{X}$ is distributed symmetrically around that value, and (ii) greater for $\phi_0 > 0$ than for $\phi_0 < 0$, despite the direction of the displacement. Concerning *the marginal effect of estimating* σ_0, the estimated control limits are $\widehat{UCL}/\widehat{LCL} = \mu_0 \pm K\frac{\hat{\sigma}_0}{\sqrt{n}C_2}$. So, (iii) the control limits are wider for $\phi_0 > 0$ than for $\phi_0 < 0$ for the same reason explained before and (iv) they will expand (contract) as σ_0 is overestimated (underestimated) with the widening (shrinking) of the control limits by a factor of $\hat{\sigma}_0$. Finally, concerning *the marginal effect of estimating* ϕ_0, the estimated control limits become $\widehat{UCL}/\widehat{LCL} = \mu_0 \pm K\frac{\sigma_0}{\sqrt{n}\hat{C}_2}$. As $\hat{C}_2$ is a decreasing function of $\hat{\phi}_0$, its reciprocal $1/\hat{C}_2$ is increasing for $\hat{\phi}_0$, and therefore, (v) overestimating ϕ_0 leads to higher values of the ARL for both positive and negative autocorrelation.

However, since different Phase I samples lead to different estimations, different control limits and, consequently, different ARLs, in order to evaluate the effect of the estimators on the conditional in-control performance of the $\bar{X}$ chart for AR(1) processes, the $AARL$ and $SDARL$ have been calculated using Algorithm 3 considering values of the Phase I sample size $m \in \{50, 100, 200, 500, 1000\}$, values of the true in-control autoregressive parameter $\phi_0 \in \{\pm 0.1, \pm 0.5, \pm 0.9\}$, a nominal $ARL_0 = 370.4$, subgroup size of n = 5, $\mu_0 = 0$, $\sigma_0 = 1$, $\delta = 0$ and $rep = 10000$. In Algorithm 3, μ_{EST} is the sample mean shown in (4.2.9), σ_{EST} is one of the estimators shown in equations (4.2.10) to (4.2.13) and ϕ_{EST} is one of the estimators shown in equations (4.2.14) to (4.2.18).

Tables 4.3.1 to 4.3.3 show the resulting $AARL_0$ and $SDARL_0$. From these measurements, we can drawn the following conclusions:

- When ϕ_0 is fixed and m increases, the $SDARL_0$ decreases and the $AARL_0$ gets closer to the desired $ARL_0 = 370.4$ for almost all cases, except for $\hat{\sigma}_{0,MR}$ which has, generally speaking, the worst performance among all the σ_0 estimators.
- When m is fixed the $SDARL_0$ and $AARL_0$ tend to increase with ϕ_0 in most of the cases. This is not

Algorithm 3 Calculation of $AARL$ and $SDARL$

Define m, n, μ_0, σ_0, ϕ_0, ARL_0, δ, rep, μ_{EST}, σ_{EST}, ϕ_{EST}.
$K \leftarrow \Phi^{-1}\left(1 - \frac{1}{2ARL_0}\right)$.
$C_2 \leftarrow C_2\left(n, \phi_0\right) = \sqrt{\frac{n}{n + 2\left(\frac{\phi_0^{n+1} - n\phi_0^2 + (n-1)\phi_0}{(\phi_0 - 1)^2}\right)}}$.
$r \leftarrow 1$.
while $r \leq rep$ **do**
Generate $X_1, X_2, \ldots, X_m$ based on the AR(1) model
$$X_j - \mu_0 - \delta\sigma_0 = \phi_0\left(X_{j-1} - \mu_0 - \delta\sigma_0\right) + \varepsilon_j, \text{ where } \varepsilon_j \sim N\left(0, \sigma_0\sqrt{1 - \phi_0^2}\right).$$
Estimate μ_0, σ_0 and ϕ_0 with μ_{EST}, σ_{EST} and ϕ_{EST} and denote them as $\hat{\mu}_{0,r}, \hat{\sigma}_{0,r}$ and $\hat{\phi}_{0,r}$, respectively.
if $\left|\hat{\phi}_{0,r}\right| < 1$ **then**
$$\hat{C}_{2,r} \leftarrow C_2\left(n, \hat{\phi}_{0,r}\right) \leftarrow \sqrt{\frac{n}{n + 2\left(\dfrac{\hat{\phi}_{0,r}^{n+1} - n\hat{\phi}_{0,r}^2 + (n-1)\hat{\phi}_{0,r}}{(\hat{\phi}_{0,r} - 1)^2}\right)}}$$
$$\beta_r \leftarrow \Phi\left[\sqrt{n}C_2\left(\frac{\hat{\mu}_{0,r} - \mu_0 - \delta\sigma_0}{\sigma_0}\right) + K\frac{\hat{\sigma}_{0,r}C_2}{\sigma_0\hat{C}_{2,r}}\right] - \Phi\left[\sqrt{n}C_2\left(\frac{\hat{\mu}_{0,r} - \mu_0 - \delta\sigma_0}{\sigma_0}\right) - K\frac{\hat{\sigma}_{0,r}C_2}{\sigma_0\hat{C}_{2,r}}\right].$$
$$ARL_r \leftarrow \frac{1}{1 - \beta_r}.$$
$r \leftarrow r + 1$.
else
Eliminate the current sample $X_1, X_2, \ldots, X_m$.
end if
end while
$AARL \leftarrow \text{mean}(ARL_1, \ldots, ARL_{rep})$.
$SDARL \leftarrow \text{stdev}(ARL_1, \ldots, ARL_{rep})$.

observed just in few cases, for instance, when ϕ_0 changes from -0.9 to -0.5.

- Considering the estimators of ϕ_0, the robust estimators $\widehat{\phi}_{0,H}$ and $\widehat{\phi}_{0,MS}$ have larger $SDARL_0$ and $AARL_0$ values in most of the cases compared to the least-squares based estimators $\hat{\phi}_{0,LS}, \hat{\phi}_{0,LS1}$ and $\hat{\phi}_{0,Q}$.
- For positive (negative) autocorrelation $\hat{\phi}_{0,LS}$ ($\hat{\phi}_{0,LS1}$) seems to have a better performance in most of the cases.
- Considering estimators of the process standard deviation, $\hat{\sigma}_{0,SQ}$ has the best performance for all combinations of m and ϕ_0 over all ϕ_0 estimators. In contrast, $\hat{\sigma}_{0,MR}$ has the worst performance, except in some particular cases (m = 50 and $\phi_0 = \pm 0.1$). This can be explained by considering that the overestimation of one or both parameters lead to higher ARL values, but this increase in the ARL is reduced when one of them is underestimated. In particular, in those cases $\hat{\sigma}_{0,MR}$ underestimates σ_0 more frequently than the other estimators.
- The values of the median of the ARL_0, $MARL_0$, (which are not reported on the tables for the sake of brevity) are below the nominal value $ARL_0 = 370.4$ for almost all cases, (without considering the results for $\hat{\sigma}_{0,MR}$).

The fact that, when m is fixed, the $AARL$ and $SDARL$ increases with ϕ_0 (except for the case when ϕ_0 changes from -0.9 to -0.5) is related to the condition $\left|\hat{\phi}_0\right| < 1$. This implies that when a series is generated, a value of $\hat{\phi}_0$ not satisfying this restriction is removed, thus creating some kind of censure on the estimations. As it will restrict the percentage of times a value of $\hat{\phi}_0$ is below the true value of $\phi_0 = -0.9$, this favours the overestimation of ϕ_0 and, therefore, this generates larger ARL values, in contrast with the other cases of $\phi_0 = \pm 0.1, \pm 0.5$.

Finally, the fact that the $MARL_0$ is below the nominal ARL_0 means that more than 50% of control charts with estimated parameters will have a worst performance than expected. In the next section, an approach to assess this problem is presented.

4.4 Guaranteed conditional in-control performance

To assess the problem that more than 50% of the $\bar{X}$ control charts for AR(1) processes will have an in-control conditional ARL lower than the nominal value, Gandy and Kvaløy's bootstrapping method is applied in order to adjust the $\bar{X}$ control limits to guarantee an in-control conditional performance with a certain probability. This approach has been proved to be effective for control charts monitoring the mean or variance of independent random variables and, also, for discrete data. The implementation of this methodology for AR(1) processes is detailed here.

Before explaining the methodology, it has to be pointed that exact methods are of course preferred over bootstrap-based methods, as those developed by Goedhart et al. (2017a), Goedhart et al. (2017b) and Faraz et al. (2018). In their papers, the authors used the fact that, for normal i.i.d. random variables, $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ to find the p.d.f. of the random variable $W = \hat{\sigma}/\sigma$ (which turns out to be a scaled-χ random variable if S^2 is replaced by $\hat{\sigma}^2$ in the previous formula). But, in the case of AR(1) processes, things turn to be more complicated: finding the distribution of $W = \dfrac{\hat{\sigma}/\hat{C}_2}{\sigma/C_2}$ is clearly much more difficult due to the presence of the term $\hat{C}_2$ and the dependence between $\bar{X}$ and S^2.

As this technique is used to guarantee an in-control conditional performance, we have to assume $\delta = 0$, which is the reduced AR(1) model stated in equation (4.2.5):

$$X_j - \mu_0 = \phi_0 \left(X_{j-1} - \mu_0\right) + \varepsilon_j,$$

Table 4.3.1: $AARL_0$ and $SDARL_0$ (below) considering $n = 5$, a target $ARL_0 = 370.4$ and $\phi_0 \in \{-0.9, -0.5\}$

m	$\hat{\phi}_0$	ϕ_0=-0.9				ϕ_0=-0.5			
		$\hat{\sigma}_{0,SQ1}$	$\hat{\sigma}_{0,UN}$	$\hat{\sigma}_{0,SQ}$	$\hat{\sigma}_{0,MR}$	$\hat{\sigma}_{0,SQ1}$	$\hat{\sigma}_{0,UN}$	$\hat{\sigma}_{0,SQ}$	$\hat{\sigma}_{0,MR}$
50	$\widehat{\phi}_{0,LS}$	544082.62	665609.84	367004.96	2.30×10^{12}	651.70	695.06	574.84	16523.11
		3.50×10^{7}	4.30×10^{7}	2.34×10^{7}	1.28×10^{14}	1565.16	1708.87	1318.68	88032.32
	$\widehat{\phi}_{0,LS1}$	400533.41	488281.51	272046.37	1.25×10^{12}	544.05	579.24	481.52	11277.06
		2.55×10^{7}	3.12×10^{7}	1.72×10^{7}	6.46×10^{13}	1327.59	1447.62	1121.34	54703.87
	$\widehat{\phi}_{0,Q}$	500072.80	611203.21	337935.40	2.25×10^{12}	586.15	624.89	517.46	14586.58
		3.24×10^{7}	3.97×10^{7}	2.17×10^{7}	1.28×10^{14}	1503.91	1642.44	1266.46	76156.97
	$\widehat{\phi}_{0,H}$	471691.92	575712.99	319652.77	1.89×10^{12}	10658.92	12157.55	8258.10	1.44×10^{6}
		2.99×10^{7}	3.66×10^{7}	2.00×10^{7}	1.02×10^{14}	225328.63	262173.56	167628.64	6.17×10^{7}
	$\widehat{\phi}_{0,MS}$	788412.55	968814.07	527222.13	1.97×10^{12}	25268.01	29413.75	18810.87	3.47×10^{6}
		5.27×10^{7}	6.50×10^{7}	3.50×10^{7}	1.02×10^{14}	765123.05	902986.01	553499.36	2.66×10^{8}
100	$\widehat{\phi}_{0,LS}$	30266.48	32878.77	25709.09	7.89×10^{11}	460.76	473.66	436.28	7435.72
		2.30×10^{6}	2.51×10^{6}	1.92×10^{6}	4.91×10^{13}	484.73	502.10	452.13	15224.79
	$\widehat{\phi}_{0,LS1}$	28039.74	30449.55	23833.51	7.64×10^{11}	420.08	431.64	398.12	6349.45
		2.13×10^{6}	2.33×10^{6}	1.79×10^{6}	4.90×10^{13}	439.20	454.74	410.02	12488.91
	$\widehat{\phi}_{0,Q}$	29954.91	32543.57	25439.19	7.85×10^{11}	431.35	443.32	408.63	6913.39
		2.29×10^{6}	2.50×10^{6}	1.92×10^{6}	4.91×10^{13}	470.07	486.89	438.50	15615.99
	$\widehat{\phi}_{0,H}$	38483.87	41911.22	32524.90	7.79×10^{11}	1487.09	1556.89	1359.07	45961.52
		3.11×10^{6}	3.41×10^{6}	2.60×10^{6}	4.91×10^{13}	30813.79	32972.05	26954.65	1.59×10^{6}
	$\widehat{\phi}_{0,MS}$	34107.96	37015.50	29028.53	1.25×10^{12}	1685.83	1767.18	1536.79	73717.61
		2.35×10^{6}	2.57×10^{6}	1.97×10^{6}	9.21×10^{13}	35721.72	38224.92	31245.21	3.73×10^{6}
200	$\widehat{\phi}_{0,LS}$	1420.45	1450.52	1362.61	2.07×10^{9}	406.21	411.51	395.88	5421.73
		12473.90	12838.63	11779.30	2.00×10^{11}	246.00	249.95	238.34	7050.76
	$\widehat{\phi}_{0,LS1}$	1357.95	1386.56	1302.94	1.87×10^{9}	387.85	392.86	378.06	5038.73
		11934.18	12281.82	11272.01	1.80×10^{11}	233.60	237.32	226.38	6365.05
	$\widehat{\phi}_{0,Q}$	1374.19	1403.19	1318.39	2.10×10^{9}	389.54	394.58	379.69	5140.19
		12212.97	12570.22	11532.64	2.03×10^{11}	241.58	245.45	234.05	7916.00
	$\widehat{\phi}_{0,H}$	1464.55	1495.65	1404.74	1.82×10^{9}	558.03	566.21	542.13	8304.94
		12458.55	12820.29	11769.39	1.76×10^{11}	864.57	881.86	831.20	19295.62
	$\widehat{\phi}_{0,MS}$	1587.06	1621.59	1520.70	1.67×10^{9}	590.79	599.95	573.01	9716.69
		14849.82	15291.67	14008.95	1.58×10^{11}	1511.04	1544.85	1446.04	47372.59
500	$\widehat{\phi}_{0,LS}$	566.14	569.45	559.61	349309.19	385.55	387.49	381.70	4682.43
		884.13	891.06	870.46	7.49×10^{6}	134.97	135.79	133.35	2633.44
	$\widehat{\phi}_{0,LS1}$	555.83	559.06	549.44	336627.52	378.46	380.36	374.69	4553.42
		867.50	874.28	854.11	7.22×10^{6}	132.26	133.06	130.68	2548.84
	$\widehat{\phi}_{0,Q}$	557.01	560.25	550.60	341083.00	378.73	380.64	374.97	4567.88
		870.39	877.21	856.96	7.40×10^{6}	133.22	134.02	131.63	2592.82
	$\widehat{\phi}_{0,H}$	573.22	576.57	566.59	326122.13	430.05	432.31	425.59	5403.52
		904.19	911.30	890.16	6.48×10^{6}	282.90	284.73	279.29	4879.89
	$\widehat{\phi}_{0,MS}$	549.88	553.09	543.53	308454.04	393.77	395.82	389.70	4831.84
		912.59	919.83	898.32	6.30×10^{6}	306.47	308.46	302.54	5541.83
1000	$\widehat{\phi}_{0,LS}$	447.85	449.04	445.49	71486.07	376.02	376.96	374.16	4395.53
		338.41	339.52	336.19	230308.55	91.14	91.41	90.60	1643.23
	$\widehat{\phi}_{0,LS1}$	443.77	444.95	441.43	70239.52	372.56	373.48	370.71	4335.59
		335.19	336.29	333.00	225920.60	90.22	90.49	89.69	1617.41
	$\widehat{\phi}_{0,Q}$	443.86	445.04	441.52	70453.98	372.49	373.41	370.65	4336.57
		335.65	336.75	333.45	227943.45	90.35	90.62	89.82	1624.70
	$\widehat{\phi}_{0,H}$	450.79	451.99	448.41	70585.32	398.53	399.54	396.51	4754.35
		344.17	345.31	341.92	217976.72	168.74	169.26	167.71	2737.34
	$\widehat{\phi}_{0,MS}$	419.82	420.93	417.63	61810.76	357.27	358.17	355.49	4076.06
		327.13	328.20	324.99	192976.31	179.09	179.64	178.00	2808.90

Table 4.3.2: $AARL_0$ and $SDARL_0$ (below) considering $n = 5$, a target $ARL_0 = 370.4$ and $\phi_0 \in \{-0.1, 0.1\}$

m	$\hat{\phi}_0$	ϕ_0=-0.1				ϕ_0=0.1			
		$\hat{\sigma}_{0,SQ1}$	$\hat{\sigma}_{0,UN}$	$\hat{\sigma}_{0,SQ}$	$\hat{\sigma}_{0,MR}$	$\hat{\sigma}_{0,SQ1}$	$\hat{\sigma}_{0,UN}$	$\hat{\sigma}_{0,SQ}$	$\hat{\sigma}_{0,MR}$
50	$\widehat{\phi}_{0,LS}$	1620.71	1789.80	1339.08	1033.70	1610.71	1763.95	1350.66	280.81
		48417.12	55796.14	36691.68	3551.36	17226.49	19499.87	13526.32	520.87
	$\widehat{\phi}_{0,LS1}$	1938.59	2157.91	1577.79	1019.33	1974.57	2172.44	1641.10	295.93
		73728.14	85349.09	55379.18	3771.50	24368.48	27700.96	18976.34	571.10
	$\widehat{\phi}_{0,Q}$	4490.92	5164.48	3430.85	1124.75	2306.18	2545.14	1904.98	321.98
		321203.66	377492.43	234220.79	4294.57	25672.65	29069.10	20141.10	816.35
	$\widehat{\phi}_{0,H}$	53200.03	62506.88	38881.33	13083.56	428797.72	523484.99	290560.40	1048.62
		2.18×10^6	2.61×10^6	1.54×10^6	430509.01	2.70×10^7	3.32×10^7	1.81×10^7	14647.19
	$\widehat{\phi}_{0,MS}$	167298.12	201444.55	116540.99	166753.75	6.50×10^7	8.44×10^7	3.91×10^7	8927.31
		9.60×10^6	1.17×10^7	6.52×10^6	1.27×10^7	6.48×10^9	8.41×10^9	3.89×10^9	665858.27
100	$\widehat{\phi}_{0,LS}$	592.30	610.80	557.40	760.16	689.31	712.56	645.63	247.77
		1337.77	1396.90	1228.25	962.55	2566.69	2694.44	2331.91	249.33
	$\widehat{\phi}_{0,LS1}$	592.99	611.59	557.91	750.46	723.25	748.01	676.79	252.95
		1406.87	1469.93	1290.19	957.78	2844.98	2988.58	2581.33	259.77
	$\widehat{\phi}_{0,Q}$	601.15	620.08	565.46	768.87	726.54	751.33	680.02	256.52
		1405.84	1468.28	1290.23	1056.27	2630.99	2760.76	2392.33	266.86
	$\widehat{\phi}_{0,H}$	1759.55	1839.88	1611.60	1545.99	2480.67	2609.69	2245.34	380.61
		24723.70	26386.40	21739.70	7469.70	47904.74	51257.49	41908.32	981.28
	$\widehat{\phi}_{0,MS}$	1894.61	1980.66	1735.90	1845.27	7913.19	8452.99	6949.69	542.16
		15348.02	16267.56	13681.09	21277.34	292278.99	315934.92	250593.93	2471.99
200	$\widehat{\phi}_{0,LS}$	448.24	454.34	436.34	667.03	475.26	481.91	462.31	233.14
		428.42	435.88	413.95	479.71	546.31	556.07	527.40	135.88
	$\widehat{\phi}_{0,LS1}$	446.12	452.19	434.28	661.90	481.86	488.63	468.68	235.16
		430.30	437.81	415.75	477.17	561.55	571.63	542.02	137.94
	$\widehat{\phi}_{0,Q}$	447.67	453.78	435.78	665.40	483.46	490.26	470.22	235.82
		436.11	443.74	421.33	486.47	567.39	577.58	547.64	140.42
	$\widehat{\phi}_{0,H}$	623.05	632.61	604.48	858.80	679.97	690.81	658.94	279.81
		1186.23	1210.57	1139.27	1202.67	1829.79	1872.04	1748.67	285.63
	$\widehat{\phi}_{0,MS}$	632.94	642.72	613.94	875.71	768.25	781.09	743.38	300.16
		1299.05	1326.33	1246.48	1344.29	2292.10	2344.55	2191.30	386.60
500	$\widehat{\phi}_{0,LS}$	399.22	401.26	395.17	630.21	402.77	404.85	398.66	228.58
		198.74	199.98	196.29	267.04	230.81	232.26	227.94	80.44
	$\widehat{\phi}_{0,LS1}$	398.14	400.17	394.10	628.13	404.37	406.45	400.23	229.29
		198.60	199.84	196.16	266.37	232.53	234.00	229.63	80.87
	$\widehat{\phi}_{0,Q}$	398.25	400.29	394.21	628.53	404.55	406.64	400.42	229.43
		198.80	200.03	196.35	267.51	232.27	233.74	229.38	81.08
	$\widehat{\phi}_{0,H}$	445.90	448.28	441.20	691.01	450.01	452.43	445.23	243.39
		350.92	353.26	346.28	463.85	382.19	384.77	377.10	127.28
	$\widehat{\phi}_{0,MS}$	457.56	460.01	452.72	711.43	452.98	455.43	448.12	242.83
		342.28	344.54	337.82	464.97	474.64	478.08	467.83	143.50
1000	$\widehat{\phi}_{0,LS}$	382.63	383.59	380.72	615.92	385.67	386.64	383.74	227.86
		126.10	126.48	125.35	175.40	147.17	147.62	146.28	54.81
	$\widehat{\phi}_{0,LS1}$	382.06	383.02	380.16	614.87	386.35	387.32	384.41	228.20
		126.04	126.42	125.28	175.18	147.66	148.11	146.77	54.96
	$\widehat{\phi}_{0,Q}$	382.11	383.07	380.20	615.00	386.45	387.42	384.51	228.25
		126.27	126.64	125.51	175.68	147.77	148.22	146.88	55.05
	$\widehat{\phi}_{0,H}$	403.16	404.19	401.11	643.85	408.50	409.55	406.41	236.22
		195.08	195.69	193.88	276.04	213.16	213.83	211.84	83.59
	$\widehat{\phi}_{0,MS}$	419.62	420.70	417.47	671.54	397.28	398.29	395.25	230.30
		202.57	203.20	201.32	289.10	219.36	220.04	217.99	88.06

Table 4.3.3: $AARL_0$ and $SDARL_0$ (below) considering $n = 5$, a target $ARL_0 = 370.4$ and $\phi_0 \in \{0.5, 0.9\}$

m	$\hat{\phi}_0$	ϕ_0=0.5				ϕ_0=0.9			
		$\hat{\sigma}_{0,SQ1}$	$\hat{\sigma}_{0,UN}$	$\hat{\sigma}_{0,SQ}$	$\hat{\sigma}_{0,MR}$	$\hat{\sigma}_{0,SQ1}$	$\hat{\sigma}_{0,UN}$	$\hat{\sigma}_{0,SQ}$	$\hat{\sigma}_{0,MR}$
50	$\widehat{\phi}_{0,LS}$	5365.59	6060.05	4238.40	26.39	7443.79	8685.45	5513.63	2.48
		129100.27	149659.02	96735.01	21.15	530830.23	626862.43	383482.71	0.43
	$\widehat{\phi}_{0,LS1}$	9545.14	10888.51	7394.38	28.99	19398.43	22918.83	14017.94	2.57
		266554.48	311566.76	196527.96	24.76	1.48×10^6	1.76×10^6	1.04×10^6	0.46
	$\widehat{\phi}_{0,Q}$	8737.04	9938.09	6806.04	29.17	7720.53	9006.20	5721.53	2.54
		224133.80	261786.89	165517.36	25.63	541278.78	639291.42	390924.97	0.46
	$\widehat{\phi}_{0,H}$	104539.29	124128.75	74804.60	33.81	6037.57	6983.52	4548.49	2.50
		4.32×10^6	5.19×10^6	3.02×10^6	51.77	317140.51	372071.32	232084.23	0.47
	$\widehat{\phi}_{0,MS}$	232015.66	280327.24	160553.72	48.47	10315.09	12108.72	7549.64	2.51
		1.72×10^7	2.10×10^7	1.16×10^7	104.98	794721.67	942422.58	569441.36	0.51
100	$\widehat{\phi}_{0,LS}$	1470.73	1546.74	1332.57	27.37	6458.96	6933.48	5616.71	2.62
		52298.50	56244.94	45292.21	13.41	345705.43	374285.38	295454.48	0.35
	$\widehat{\phi}_{0,LS1}$	1868.82	1972.18	1681.92	28.61	10036.27	10807.80	8672.67	2.66
		76331.58	82253.56	65848.08	14.38	565547.39	613831.10	480961.21	0.37
	$\widehat{\phi}_{0,Q}$	1898.77	2004.35	1707.94	28.72	6708.03	7196.49	5840.25	2.66
		78954.90	85098.88	68081.72	15.04	330970.33	358150.15	283146.66	0.37
	$\widehat{\phi}_{0,H}$	2618.93	2753.76	2372.62	30.90	7667.98	8242.26	6650.30	2.63
		38524.36	41173.93	33780.73	23.87	388362.97	420365.84	332069.84	0.38
	$\widehat{\phi}_{0,MS}$	6943.32	7365.58	6181.07	38.22	10218.40	11035.17	8781.08	2.65
		129974.52	139624.53	112819.49	44.23	734800.56	799019.30	622591.35	0.42
200	$\widehat{\phi}_{0,LS}$	553.08	561.43	536.85	28.41	910.56	929.40	874.28	2.73
		1042.06	1063.00	1001.66	9.58	7128.67	7322.25	6758.90	0.27
	$\widehat{\phi}_{0,LS1}$	584.11	593.03	566.77	29.03	1031.31	1053.17	989.26	2.75
		1129.55	1152.50	1085.26	9.91	8398.21	8630.17	7955.41	0.27
	$\widehat{\phi}_{0,Q}$	585.41	594.36	568.03	29.07	1016.84	1038.35	975.48	2.75
		1126.42	1149.26	1082.36	10.04	8329.52	8559.64	7890.25	0.27
	$\widehat{\phi}_{0,H}$	780.31	793.67	754.44	30.05	948.47	968.10	910.67	2.74
		3040.51	3116.44	2895.07	14.13	6784.37	6964.56	6439.85	0.28
	$\widehat{\phi}_{0,MS}$	1275.04	1300.28	1226.38	34.47	1164.62	1190.30	1115.31	2.78
		9941.34	10236.10	9380.47	21.95	12465.46	12833.56	11764.74	0.33
500	$\widehat{\phi}_{0,LS}$	432.24	434.54	427.69	29.04	479.18	482.02	473.57	2.83
		361.39	363.82	356.60	6.01	959.05	966.60	944.16	0.16
	$\widehat{\phi}_{0,LS1}$	440.12	442.47	435.47	29.29	496.25	499.21	490.40	2.84
		370.66	373.16	365.72	6.09	1005.53	1013.51	989.81	0.17
	$\widehat{\phi}_{0,Q}$	440.16	442.51	435.51	29.29	495.83	498.79	489.98	2.84
		369.04	371.52	364.14	6.11	1002.15	1010.08	986.50	0.17
	$\widehat{\phi}_{0,H}$	482.46	485.13	477.18	29.80	494.76	497.72	488.92	2.83
		544.45	548.40	536.67	8.22	1017.76	1025.88	1001.75	0.17
	$\widehat{\phi}_{0,MS}$	582.91	586.27	576.26	32.49	600.19	603.92	592.82	2.90
		749.98	755.67	738.74	10.92	1314.75	1325.57	1293.44	0.21
1000	$\widehat{\phi}_{0,LS}$	396.75	397.76	394.74	29.28	423.14	424.28	420.86	2.87
		199.67	200.29	198.44	4.24	466.05	467.67	462.84	0.11
	$\widehat{\phi}_{0,LS1}$	400.12	401.14	398.08	29.40	429.90	431.06	427.58	2.88
		201.94	202.56	200.69	4.27	475.80	477.46	472.51	0.11
	$\widehat{\phi}_{0,Q}$	400.08	401.10	398.04	29.40	429.86	431.03	427.55	2.88
		202.16	202.79	200.92	4.28	475.82	477.47	472.53	0.11
	$\widehat{\phi}_{0,H}$	416.29	417.37	414.14	29.64	430.67	431.84	428.34	2.87
		247.68	248.46	246.13	5.63	488.93	490.64	485.53	0.12
	$\widehat{\phi}_{0,MS}$	487.82	489.13	485.22	31.97	524.19	525.68	521.23	2.94
		348.86	350.03	346.53	7.49	667.39	669.84	662.51	0.15

for $j = 1, 2, \ldots, m,$ where $X_0 \sim N(\mu_0, \sigma_0), \varepsilon_j \sim N(0, \sigma_\varepsilon)$ and $|\phi_0| < 1$. At this stage, it is necessary to introduce an estimator for σ_ε which was not introduced before as its estimation was not involved on the calculation of the ARL. However, in this stage, it plays a role on the calculation of the adjusted control limits, and an estimator for σ_ε is now needed. Considering the relationship shown in (4.2.2), a natural estimator for σ_ε is given by

$$\hat{\sigma}_\varepsilon = \hat{\sigma}_0 \sqrt{\left(1 - \hat{\phi}_0^2\right)}. \tag{4.4.1}$$

In order to apply the bootstrap method, let P denote the process shown in equation (4.2.5) and $\boldsymbol{\xi} = \boldsymbol{\xi}(P)$ denote the process P parameters, that is to say, $\boldsymbol{\xi} = \langle \mu_0, \sigma_0, \phi_0, \sigma_\varepsilon \rangle$. Thus, $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}(\hat{P}) = \left\langle \hat{\mu}_0, \hat{\sigma}_0, \hat{\phi}_0, \hat{\sigma}_\varepsilon \right\rangle$ stands for the estimated process parameters. The variable analyzed in the bootstrap procedure is the value of K required to get the target ARL_0 and it is computed by using Algorithm 4. In addition, due to the fact that the estimated model might differ from sample to sample, the bootstrapping methodology was applied several times and the K-values reported in Table 4.4.1 are the averaged adjusted K-values. To do so, Algorithm 4 was applied using $rep = 100$ runs, $B = 1000$ bootstrap replications, $\alpha^* = 0.9, ARL_0 = 370.4, \mu_{EST} = \bar{X}, \sigma_{EST} = \hat{\sigma}_{0,SQ}$, ϕ_{EST} is one of the estimators in equations (4.2.14) to (4.2.18) and $\sigma_{\varepsilon EST}$ is obtained using equation (4.4.1). We considered $rep = 100$ since this value provides a good balance between computational time and estimation accuracy. For instance, the relative standard error of the estimation of the averaged adjusted K-value for $\hat{\phi}_{0,LS}$ when $m = 50$ and $\phi_0 = 0.9$ is 1.76%, being this the highest value for $\hat{\phi}_{0,LS}$ with other combinations of m and ϕ_0. Besides the adjusted K values, the 10-th quantile ($ARL_{0,0.1}$) of the distribution of the ARL_0 with adjusted control limits is also reported based on 10000 ARL values, calculated with Algorithm 3 but using these adjusted K-values.

From Table 4.4.1 we could see that the control limits become thinner as m increases. This happens because the estimators tend to be more accurate. In almost all cases the $ARL_{0,0.1}$ is within 10% the nominal ARL_0 value except for the cases where $\phi_0 = \pm 0.9$ and for small values of m. However, the performance is better than the non-adjusted case (not shown here). Even though an in-control ARL_0 could be guaranteed by using this methodology, it is well known that the adjustment of the control limits has a negative effect on the out-of-control ARL. In order to explore that effect, the ARL_1 of the $\bar{X}$ control chart has been calculated considering Algorithm 3 with $\delta = 1$, by using the K-value as in Algorithm 3 for the unadjusted (U) control limits case and by replacing that value with the one reported on Table 4.4.1 for the case of adjusted (A) control limits.

Table 4.4.2 shows the values of $ARL_{1,0.9}$, i.e., the 90-th quantile of the distribution of the ARL_1 which means that around 90% of the control charts will have an ARL_1 of at most that value. Furthermore, the ARL_1 values for a shift size $\delta = 1$ and *known* parameters are shown in the last row of the table. It can be seen that the increment in the ARL_1 due to the adjustment is lower for $\phi_0 < 0$ than for $\phi_0 > 0$, and in almost all cases $\hat{\phi}_{0,LS}$ has the lowest $ARL_{1,0.9}$ among the ϕ_0 estimators considered here. As ϕ_0 increases, a larger retrospective Phase I dataset of observations is required to reduce the loss in statistical sensitivity. In fact, it can be seen that $\phi_0 = 0.9$ is the only case where the $ARL_{1,0.9}$ does not decrease as m increases. This is observed because, for the generated series, the constraint $\left|\hat{\phi}_0\right| < 1$ restricts the percentage of times a value $\hat{\phi}_0$ is above $\phi_0 = 0.9$, especially for small values of m, where the estimators accuracy is not good enough to mitigate the impact of the censure.

4.5 Illustrative example

In this Section, the sampling methodology is illustrated using a dataset found in Box et al. (2015) corresponding to viscosity readings (unit not specified), taken every hour from a chemical process and labeled as Series D. Before setting up the $\bar{X}$ control chart, recall that the Phase I is done by sampling m consecutive individual $(n = 1)$ observations, collected to estimate the process parameters and to check for process stability. After that, the Phase II process monitoring is performed by collecting, each h hours, independent samples of n consecutive autocorrelated observations.

Algorithm 4 Calculation of the adjusted K-values

Define m, n, μ_0, σ_0, ϕ_0, ARL_0, $\alpha^*, B, rep, \mu_{EST}, \sigma_{EST}, \phi_{EST}, \sigma_{\varepsilon EST}$.
$\delta \leftarrow 0$.
$\alpha \leftarrow \frac{1}{ARL_0}$.
$r \leftarrow 1$.
while $r \leq rep$ **do**
  Generate $X_1, X_2, \ldots, X_m$ based on the AR(1) model:
$$X_j - \mu_0 = \phi_0(X_{j-1} - \mu_0) + \varepsilon_j, \text{ where } \varepsilon_j \sim N\left(0, \sigma_0\sqrt{1-\phi_0^2}\right)$$
  Estimate $\mu_0, \sigma_0, \phi_0, \sigma_\varepsilon$ with $\mu_{EST}, \sigma_{EST}, \phi_{EST}, \sigma_{\varepsilon EST}$ and denote them as $\hat{\mu}_{0,r}, \hat{\sigma}_{0,r}, \hat{\phi}_{0,r}$ and $\hat{\sigma}_{\varepsilon,r}$, respectively.
  if $\left|\hat{\phi}_{0,r}\right| < 1$ **then**
$$\hat{C}_{2,r} \leftarrow C_2\left(n, \hat{\phi}_{0,r}\right) = \sqrt{\frac{n}{n + 2\left(\frac{\hat{\phi}_{0,r}^{n+1} - n\hat{\phi}_{0,r}^2 + (n-1)\hat{\phi}_{0,r}}{(\hat{\phi}_{0,r}-1)^2}\right)}}.$$
    $b \leftarrow 1$.
    while $b \leq B$ **do**
      Generate $X_1^*, X_2^*, \ldots, X_m^*$ based on the estimated AR(1) model:
$$X_j - \hat{\mu}_{0,r} = \hat{\phi}_{0,r}\left(X_{j-1} - \hat{\mu}_{0,r}\right) + \varepsilon_j, \text{ where } \varepsilon_j \sim N\left(0, \hat{\sigma}_{\varepsilon,r}\right)$$
      Estimate $\mu_0, \sigma_0, \phi_0, \sigma_\varepsilon$ with $\mu_{EST}, \sigma_{EST}, \phi_{EST}, \sigma_{\varepsilon EST}$ and denote them as $\hat{\mu}_{0,b}^*, \hat{\sigma}_{0,b}^*, \hat{\phi}_{0,b}^*$ and $\hat{\sigma}_{\varepsilon,b}^*$, respectively.
      if $\left|\hat{\phi}_{0,b}^*\right| < 1$ **then**
$$\hat{C}_{2,b}^* \leftarrow C_2\left(n, \hat{\phi}_{0,b}^*\right) = \sqrt{\frac{n}{n + 2\left(\frac{\hat{\phi}_{0,b}^{*n+1} - n\hat{\phi}_{0,b}^{*2} + (n-1)\hat{\phi}_{0,b}^*}{(\hat{\phi}_{0,b}^*-1)^2}\right)}}.$$
        $K\left(\hat{P}, \hat{\phi}_{0,b}^*\right) \leftarrow$ calculate as the solution of equation
$$1 - \alpha = \Phi\left[\sqrt{n}\hat{C}_{2,r}\left(\frac{\hat{\mu}_{0,b}^* - \hat{\mu}_{0,r}}{\hat{\sigma}_{0,r}}\right) + K\left(\hat{P}, \hat{\phi}_{0,b}^*\right)\frac{\hat{\sigma}_{0,b}^*\hat{C}_{2,r}}{\hat{\sigma}_{0,r}\hat{C}_{2,b}^*}\right] - \Phi\left[\sqrt{n}\hat{C}_{2,r}\left(\frac{\hat{\mu}_{0,b}^* - \hat{\mu}_{0,r}}{\hat{\sigma}_{0,r}}\right) - K\left(\hat{P}, \hat{\phi}_{0,b}^*\right)\frac{\hat{\sigma}_{0,b}^*\hat{C}_{2,r}}{\hat{\sigma}_{0,r}\hat{C}_{2,b}^*}\right].$$
        $b \leftarrow b + 1$.
      else
        Eliminate the current sample $X_1^*, X_2^*, \ldots, X_m^*$.
      end if
    end while
    $K_r \leftarrow$ is the α^*-th quantile of $K\left(\hat{P}, \hat{\phi}_{0,b}^*\right)$.
    $r \leftarrow r + 1$.
  else
    Eliminate the current sample $X_1, X_2, \ldots, X_m$.
  end if
end while
$K \leftarrow \text{mean}\left(K_1, \ldots, K_{rep}\right)$.

Table 4.4.1: The averaged adjusted K values for the control limits for different values of m and ϕ_0 for a desired $ARL_0 = 370.4$ and $n = 5$. $ARL_{0,0.1}$ is the 10-th quantile of the simulated ARL_0 obtained using these values.

m	$\hat{\phi}_0$	$\phi_0 = -0.9$		$\phi_0 = -0.5$		$\phi_0 = -0.1$		$\phi_0 = 0.1$		$\phi_0 = 0.5$		$\phi_0 = 0.9$	
		K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$	K	$ARL_{0,0.1}$
50	$\widehat{\phi}_{0,LS}$	4.04	161.02	3.57	356.60	3.74	312.61	3.87	317.62	4.22	292.12	5.65	94.50
	$\widehat{\phi}_{0,LS1}$	4.79	648.88	3.63	353.40	3.76	306.68	3.88	320.34	4.20	311.14	7.09	1111.58
	$\widehat{\phi}_{0,Q}$	4.53	378.96	3.66	348.56	3.78	311.12	3.90	314.71	4.22	315.94	6.82	647.10
	$\widehat{\phi}_{0,H}$	4.59	469.62	3.91	355.03	4.03	307.54	4.16	326.62	4.47	335.12	7.57	1783.45
	$\widehat{\phi}_{0,MS}$	5.09	1263.73	4.06	335.24	4.13	298.79	4.22	327.04	4.67	534.80	8.50	9279.41
100	$\widehat{\phi}_{0,LS}$	3.76	228.96	3.37	360.73	3.49	348.17	3.55	327.38	3.78	328.51	4.82	216.67
	$\widehat{\phi}_{0,LS1}$	3.95	318.46	3.40	354.39	3.50	349.12	3.56	330.61	3.77	335.01	5.04	375.16
	$\widehat{\phi}_{0,Q}$	3.96	331.29	3.40	355.35	3.51	347.95	3.56	333.90	3.77	340.25	5.15	452.91
	$\widehat{\phi}_{0,H}$	4.07	459.42	3.58	345.39	3.69	348.03	3.73	341.14	3.92	350.80	5.29	529.53
	$\widehat{\phi}_{0,MS}$	4.52	1370.08	3.71	333.95	3.74	347.52	3.74	359.39	3.94	388.82	5.78	1618.68
200	$\widehat{\phi}_{0,LS}$	3.61	331.22	3.24	357.96	3.33	358.52	3.37	354.18	3.51	346.56	4.13	285.22
	$\widehat{\phi}_{0,LS1}$	3.67	371.09	3.25	358.35	3.33	359.92	3.37	354.51	3.50	344.68	4.18	346.10
	$\widehat{\phi}_{0,Q}$	3.72	432.15	3.26	354.49	3.33	356.86	3.37	355.52	3.50	351.26	4.59	966.55
	$\widehat{\phi}_{0,H}$	3.72	440.05	3.39	367.58	3.45	350.64	3.48	354.89	3.58	346.05	4.27	398.18
	$\widehat{\phi}_{0,MS}$	4.16	1413.26	3.50	351.65	3.48	355.62	3.49	356.51	3.59	359.73	4.75	1458.54
500	$\widehat{\phi}_{0,LS}$	3.38	387.28	3.14	369.42	3.19	363.53	3.22	358.55	3.30	366.81	3.63	346.14
	$\widehat{\phi}_{0,LS1}$	3.40	399.12	3.15	371.09	3.19	360.15	3.22	360.49	3.29	365.35	3.63	359.14
	$\widehat{\phi}_{0,Q}$	3.40	397.50	3.15	368.67	3.20	364.68	3.22	358.55	3.29	361.99	3.63	365.32
	$\widehat{\phi}_{0,H}$	3.41	411.96	3.23	366.90	3.27	363.34	3.29	354.42	3.34	362.33	3.65	355.13
	$\widehat{\phi}_{0,MS}$	3.61	674.63	3.32	369.13	3.27	365.82	3.29	349.06	3.32	372.61	3.84	678.22
1000	$\widehat{\phi}_{0,LS}$	3.25	371.25	3.10	364.64	3.13	366.08	3.15	366.16	3.20	367.67	3.40	367.35
	$\widehat{\phi}_{0,LS1}$	3.26	374.95	3.10	365.92	3.13	365.60	3.15	368.23	3.20	368.50	3.41	373.20
	$\widehat{\phi}_{0,Q}$	3.26	374.01	3.10	364.67	3.13	366.50	3.15	367.81	3.20	367.37	3.41	373.21
	$\widehat{\phi}_{0,H}$	3.26	375.96	3.16	366.80	3.18	365.27	3.20	365.88	3.23	367.35	3.41	367.10
	$\widehat{\phi}_{0,MS}$	3.41	532.31	3.23	367.22	3.18	368.05	3.21	356.26	3.21	367.77	3.47	489.41

As seen in Table 4.5.1, the first 72 observations of the original series were used as a Phase I sample. They were analyzed following the Box-Jenkins approach. Autocorrelation and partial autocorrelation functions (ACF and PACF, respectively) indicated that an AR(1) model might be appropriate to fit the series. The AR(1) model was fitted: the parameters were estimated as $\hat{\mu}_0 = 8.5153$, $\hat{\sigma}_{0,SQ} = 0.4377$, and $\hat{\phi}_{0,LS} = 0.8243$. The fit was evaluated using the Ljung-Box Q (LBQ) statistics over residuals, and showed to be not significant at $\alpha = 0.05$ even at lag 48, meaning that the residuals of the fitted model are actually independent. Correlograms from the ACF and PACF over residuals agreed with this conclusion using the same α level over each lag. In addition, the Anderson-Darling goodness of fit test applied to the same residuals showed a p-value of 0.065, failing to reject the normality assumption. With this, it was concluded that the AR(1) was a suitable model for the series. Finally, in order to check for process stability, a Phase I X control chart for the AR(1) process was set considering $n = 1$. Using $\hat{C}_2 = 1$ when $n = 1$, and a value of $K = 3$ to get an expected ARL_0 of 370.4, the corresponding control limits were calculated as $\widehat{LCL} = 7.2022$ and $\widehat{UCL} = 9.8283$. This Phase I control chart is presented in Figure 4.5.1a. It shows that the process is stable and the process parameter estimates can be used to set a Phase II, online monitoring.

To show how the monitoring is done over batched samples, Phase II samples of $n = 5$ observations were collected every 24 hours, starting at the 73rd observation of the original series. This is, the first batch corresponds to observations 73 to 77, the second batch corresponds to observations 97 to 101, and so on. The data in Table 4.5.2 present the collected samples. Their corresponding means were plotted in a Phase II $\bar{X}$ control chart. In this case, the unadjusted estimated control limits are $\widehat{LCL}_U = 7.3755$ and $\widehat{UCL}_U = 9.6550$, since $\hat{C}_2 = 0.5152$ when $n = 5$ and $\hat{\phi}_{0,LS} = 0.8243$. Figure 4.5.1b shows the Phase II $\bar{X}$ control chart with these unadjusted control limits in solid lines, where two signals are triggered at samples #3 and #9, suggesting that assignable causes of variation might be found and eliminated. However, if the bootstrapping methodology is applied, then the obtained adjusted K-value (by applying Algorithm 4 with $rep = 100$ runs, $B = 1000$ bootstrap replications, $\alpha^* = 0.9$, $ARL_0 = 370.4$, $\mu_{EST} = \hat{\mu}_0 = 8.5153$, $\sigma_{EST} = \hat{\sigma}_{0,SQ} = 0.4377$, $\phi_{EST} = \hat{\phi}_{0,LS} = 0.8243$ and $\sigma_{\varepsilon EST} = \hat{\sigma}_{0,SQ}\sqrt{1 - \hat{\phi}^2_{0,LS}} = 0.2478$) is $K = 4.8633$.

Table 4.4.2: The 90-th quantile of the simulated ARL_1 values ($ARL_{1,0.9}$) considering a shift of $\delta = 1$, with $n = 5$, $ARL_0 = 370.4$

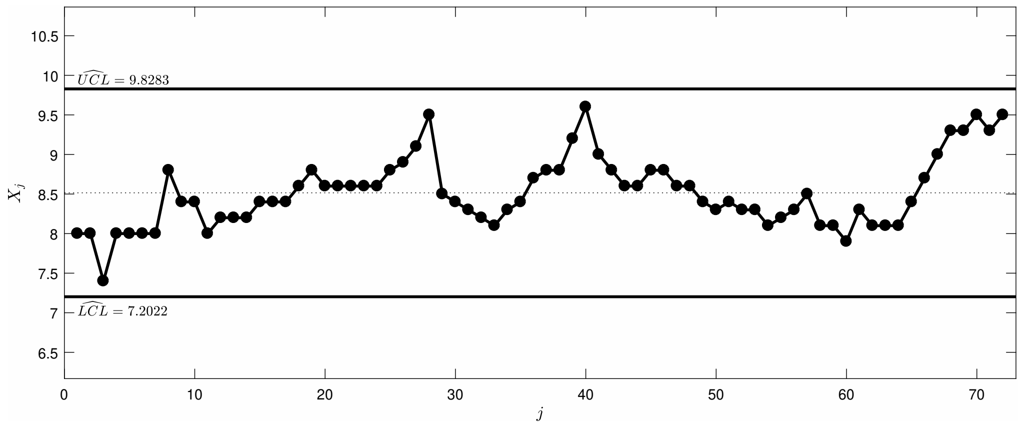
m	$\hat{\phi}$	$\phi_0 = -0.9$		$\phi_0 = -0.5$		$\phi_0 = -0.1$		$\phi_0 = 0.1$		$\phi_0 = 0.5$		$\phi_0 = 0.9$	
		U	A	U	A	U	A	U	A	U	A	U	A
50	$\widehat{\phi}_{0,LS}$	1.11	1.89	2.25	4.58	9.20	54.20	19.58	246.75	88.49	9993.36	65.84	1.27×10^{6}
	$\widehat{\phi}_{0,LS1}$	1.10	4.34	2.16	4.64	9.20	57.33	20.50	276.10	106.28	13337.03	82.81	4.88×10^{9}
	$\widehat{\phi}_{0,Q}$	1.10	3.18	2.19	4.96	9.36	62.18	20.76	289.53	105.40	14368.49	73.95	7.93×10^{8}
	$\widehat{\phi}_{0,H}$	1.12	3.79	2.81	14.20	14.55	348.63	30.80	1961.74	123.21	84292.04	67.50	1.86×10^{10}
	$\widehat{\phi}_{0,MS}$	1.13	11.32	2.81	21.11	15.23	596.88	34.40	3489.78	177.64	8.08×10^{5}	66.40	5.44×10^{11}
100	$\widehat{\phi}_{0,LS}$	1.11	1.52	1.95	2.79	6.88	18.29	13.20	49.74	54.24	701.40	116.89	1.57×10^{5}
	$\widehat{\phi}_{0,LS1}$	1.10	1.71	1.90	2.79	6.87	18.78	13.42	51.73	58.78	757.09	133.08	6.85×10^{5}
	$\widehat{\phi}_{0,Q}$	1.10	1.74	1.91	2.83	6.86	18.72	13.62	53.03	58.70	755.01	127.93	1.14×10^{6}
	$\widehat{\phi}_{0,H}$	1.11	1.99	2.22	4.68	8.78	44.90	17.21	129.34	69.45	1880.80	123.13	2.04×10^{6}
	$\widehat{\phi}_{0,MS}$	1.12	3.59	2.24	6.01	8.80	52.40	19.34	171.29	91.85	3453.44	120.70	3.68×10^{7}
200	$\widehat{\phi}_{0,LS}$	1.09	1.33	1.77	2.15	5.54	9.67	10.51	22.99	37.34	157.59	138.14	9490.33
	$\widehat{\phi}_{0,LS1}$	1.09	1.36	1.75	2.15	5.53	9.77	10.59	23.22	38.66	158.37	147.65	13927.34
	$\widehat{\phi}_{0,Q}$	1.09	1.40	1.76	2.16	5.53	9.74	10.60	23.42	38.73	162.04	147.93	1.05×10^{5}
	$\widehat{\phi}_{0,H}$	1.09	1.41	1.93	2.85	6.41	15.38	12.42	38.95	43.74	256.22	143.85	19845.73
	$\widehat{\phi}_{0,MS}$	1.09	1.98	1.93	3.25	6.43	16.50	12.59	39.88	54.80	359.77	156.24	2.54×10^{5}
500	$\widehat{\phi}_{0,LS}$	1.07	1.15	1.66	1.84	4.68	6.28	8.22	12.37	26.88	55.99	101.49	781.73
	$\widehat{\phi}_{0,LS1}$	1.07	1.16	1.66	1.84	4.68	6.26	8.23	12.43	27.22	55.93	104.16	823.70
	$\widehat{\phi}_{0,Q}$	1.07	1.16	1.66	1.84	4.68	6.29	8.27	12.44	27.25	55.69	104.15	840.95
	$\widehat{\phi}_{0,H}$	1.07	1.16	1.75	2.11	5.07	7.84	8.99	15.82	29.31	70.18	104.74	885.80
	$\widehat{\phi}_{0,MS}$	1.07	1.24	1.73	2.26	5.13	8.05	9.04	15.98	33.55	79.28	118.66	2282.66
1000	$\widehat{\phi}_{0,LS}$	1.05	1.10	1.61	1.72	4.29	5.17	7.36	9.54	22.60	35.98	79.56	265.87
	$\widehat{\phi}_{0,LS1}$	1.05	1.10	1.61	1.72	4.29	5.17	7.37	9.58	22.73	36.07	80.50	270.69
	$\widehat{\phi}_{0,Q}$	1.05	1.10	1.61	1.72	4.29	5.17	7.37	9.59	22.80	36.13	80.56	271.50
	$\widehat{\phi}_{0,H}$	1.05	1.10	1.66	1.87	4.53	5.96	7.86	11.25	23.88	41.44	80.39	275.30
	$\widehat{\phi}_{0,MS}$	1.05	1.13	1.64	1.95	4.61	6.06	7.78	11.28	26.78	44.54	94.23	412.15
Known case values		1.0321		1.5001		3.5440		5.7199		14.9949		36.1217	

Table 4.5.1: Phase I data for the illustrative example.

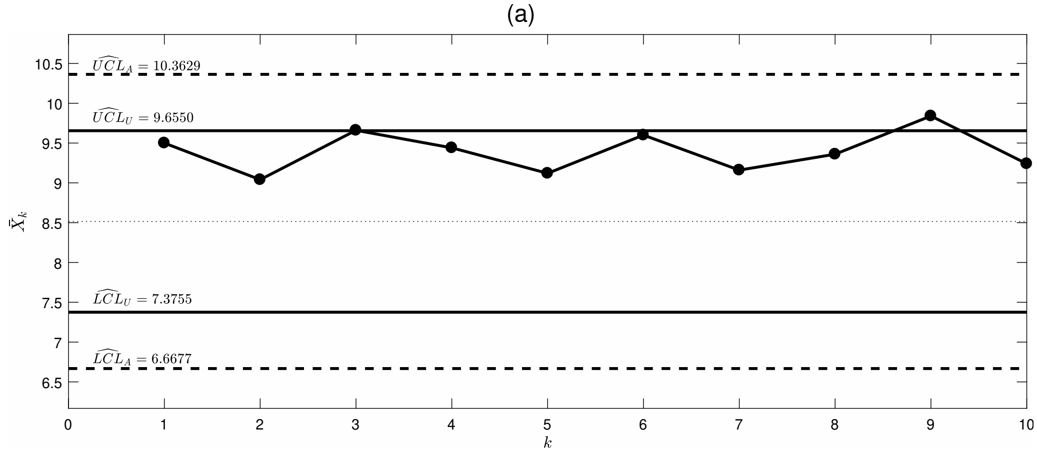
1-12	8.0	8.0	7.4	8.0	8.0	8.0	8.0	8.8	8.4	8.4	8.0	8.2
13-24	8.2	8.2	8.4	8.4	8.4	8.6	8.8	8.6	8.6	8.6	8.6	8.6
25-36	8.8	8.9	9.1	9.5	8.5	8.4	8.3	8.2	8.1	8.3	8.4	8.7
37-48	8.8	8.8	9.2	9.6	9.0	8.8	8.6	8.6	8.8	8.8	8.6	8.6
49-60	8.4	8.3	8.4	8.3	8.3	8.1	8.2	8.3	8.5	8.1	8.1	7.9
61-72	8.3	8.1	8.1	8.1	8.4	8.7	9.0	9.3	9.3	9.5	9.3	9.5

Table 4.5.2: Phase II data for the illustrative example.

k	X_1	X_2	X_3	X_4	X_5	$\bar{X}_k$
1	9.5	9.5	9.5	9.5	9.5	**9.50**
2	9.4	9.0	9.0	8.8	9.0	**9.04**
3	9.5	9.5	9.5	9.9	9.9	**9.66**
4	9.4	9.4	9.4	9.4	9.6	**9.44**
5	9.4	9.8	8.8	8.8	8.8	**9.12**
6	10.0	10.0	9.6	9.2	9.2	**9.60**
7	8.6	9.0	9.4	9.4	9.4	**9.16**
8	9.0	9.4	9.4	9.4	9.6	**9.36**
9	10.4	10.4	9.8	9.0	9.6	**9.84**
10	10.0	9.6	9.0	9.0	8.6	**9.24**



(a)



(b)

Figure 4.5.1: Control charts for viscosity data. Panel (a) corresponds to the Phase I X control chart. Panel (b) is the Phase II $\bar{X}$ control chart with unadjusted limits (solid) and adjusted limits (dashed).

With this value, the adjusted control limits are given by $\widehat{LCL}_A = 6.6677$ and $\widehat{UCL}_A = 10.3629$. These adjusted limits are shown in Figure 4.5.1b in dashed lines. In this case, no signal is triggered by the chart.

4.6 Conclusions and future work

It is well known that control charts underperform in the presence of autocorrelation, or when they are implemented with estimated parameters. The $\bar{X}$ control chart performs worst for positive autocorrelation than for negative one, since there is a widening (shrinking) on the control limits for $\phi_0 > 0$ ($\phi_0 < 0$) due to the coefficient C_2, even in the *known* parameters case. With this in mind, the relationships between the conditional ARL and the estimated values of μ_0, σ_0 and ϕ_0 for the $\bar{X}$ control chart monitoring stationary AR(1) processes have been presented considering two of the estimators fixed at their corresponding true parameter values. This gives an insight about the dynamics between the ARL and the estimation of the parameters. The results presented in this paper have shown that (i) the overestimation of ϕ_0 and/or σ_0 leads to higher ARL values; (ii) the effect on the estimation of μ_0 is symmetrical around the current process mean $\mu_0 + \delta\sigma_0$ and leads to lower ARL values despite it is over or underestimated. Concerning the estimators considered in this study, $\hat{\sigma}_{0,SQ}$ has the best performance over the other σ_0 estimators whereas $\hat{\phi}_{0,LS}$ is suggested to be used for ϕ_0, since it provides a good performance in both in-control and out-of-control situations. It has also been found that under estimated parameters, more than 50% of the times the ARL_0 of the control chart is below the nominal value $ARL_0 = 370.4$.

For the control limits adjustment, Gandy and Kvaløy's bootstrap methodology was applied having the desired effect on the ARL_0 for almost all cases, except for those with high levels of autocorrelation ($\phi_0 = \pm 0.9$). It is also worth to note that the effect on the ARL_1 is not severe for cases $\phi_0 < 0$, but it increases with ϕ_0. This effect can be mitigated by collecting larger Phase I retrospective samples, but for $\phi_0 = 0.9$ it still requires at least 1000 observations.

The evaluation of control charts for monitoring the variance of autoregressive processes is left as a future work, as well as a comparison with the EWMA and CUSUM control charts for this kind of processes assuming the definition of 'better performance' provided by Zwetsloot and Woodall (2017). The adjustment of the control chart limits via exact or bootstrap-based methods ensures a conditional in-control performance with a certain probability, for both cases of independent and autocorrelated data. However, there is always an increase on the out-of-control ARL that it is worth to consider and that has not been thoroughly studied even for the i.i.d. case. With this in mind and the lack of knowledge about the distribution of the random variable $W = \dfrac{\hat{\sigma}/\hat{C}_2}{\sigma/C_2}$, this issue is left as a future work.

Finally, the extension of the $\bar{X}$ control chart for AR(p) processes with $p > 1$ is also left as a future work, since to obtain the control limits, we need to calculate the variance of $\bar{X}$, which is given by

$$Var\left[\bar{X}\right] = \frac{\sigma_0^2}{n}\left[1 + 2\sum_{i=1}^{n-1}\left(1 - \frac{i}{n}\right)\rho_i\right],$$

where ρ_i has to be computed according to the type of process under study. For AR(p) processes, the ρ_i's have to be calculated recursively according to the Yule-Walker equations

$$\rho_i = \phi_1\rho_{i-1} + \phi_2\rho_{i-2} + \ldots + \phi_p\rho_{i-p}, \text{ for } i > 0.$$

which clearly might complicate the calculations in opposition of the case of $p = 1$, where $\rho_i = \phi^i$, for all i.

Chapter 5. The conditional performance of the modified S^2 control chart for AR(1) processes with estimated variance

Abstract

Control charts are widely used tools in Statistical Process Monitoring (SPM) because they are powerful in detecting departures from in-control situations. However, their power relies on the validation of the assumptions made on their design, among which are the prior knowledge of the in-control process parameters and uncorrelated (or independent) data, while in practice parameters are often estimated and several processes deal with streams of autocorrelated data, such as chemical processes. The evaluation of control charts performance when dealing with estimated parameters has recently been done by means of the Standard Deviation of the ARL, $SDARL$, as this performance measure stands for the sampling variability of the ARL due to the Phase I estimation, also called the “practitioner-to-practitioner” variation. In this study, the performance of the modified S^2 control chart for AR(1) processes is evaluated assuming unknown process variance and using the $SDARL$ as the performance measure. A bootstrapping technique to adjust the chart control limit is developed to have a guaranteed in-control conditional performance, with a certain probability, and the effect on the out-of-control ARL is studied by simulation. Results show that the overestimation of the variance has a strong effect on both in-control and out-of-control ARLs.

Keywords: AR(1) process; modified S^2 control chart; variance monitoring; estimated parameters; $SDARL$.

5.1 Introduction

When managing a process or a system, departures from a desired in-control or target state are wanted to be avoided and, when that occurs, to be detected as quickly as possible. The identification of these situations can be made via control charts, well known tools used in Statistical Process Monitoring SPM due to their usefulness and their easiness to be understood and implemented by practitioners. Nevertheless, their detection capability relies on the validation of the assumptions made at the design stage and many of them are scarcely met in practice, leading to unexpected and/or unpredictable charts performances.

One of the common assumption made on charts' design is that data have to be independent. Even though there are several applications were this is suitable, there are also several applications were it is not. For instance, Montgomery (2007) provided several examples of the implementation of control charts for autocorrelated processes using data taken from the chemical industry. Furthermore, a significant level of autocorrelation is expected to be present when the measurements are taken from the same object, when the data are collected over small periods of time, or when dealing with continuous flows of data. Concerning autocorrelated data, two literature reviews were found: Psarakis and Papaleonida (2007) which considered Statistical Process Control, SPC, techniques whereas Prajapati and Singh (2012) just focused on control charts.

The monitoring of autocorrelated data using control charts is usually done by means of: (i) the modification of the traditional control charts by taking into account the process true variance due to the presence of autocorrelation when setting the control limits, (such charts are known as *modified* control charts), (ii) the time series approach, where the residuals of a fitted time series model are monitored, (such charts are known as *residuals* control charts), or (iii) another approach from which a control chart with an specific name is designed. For instance, Vasilopoulos and Stamboulis (1978) and Amin et al. (1997) modified the $\bar{X}$ and S^2 control charts for monitoring the mean and the variance, respectively, of AR(1) processes. Concerning *residuals* control charts, Dawod et al. (2017) compared the Shewhart, EWMA and CUSUM control charts applied to the residuals of AR(1), MA(1) and ARMA(1,1) processes, following the work of Runger and Willemain (1995) and Snoussi et al. (2005) where the process parameters are supposed to be accurately estimated. Besides modified and residuals control charts, Alshraideh and Khatatbeh (2014) proposed the Gaussian Process control charts for the monitoring of the mean; Zhang and Pintar (2015) the Exponentially Weighted Mean Square (EWMS) control chart for variance monitoring; Harris et al. (2016) proposed a multivariate control chart, applied to autocorrelated tool wear processes; Osei-Aning et al. (2017a) designed mixed CUSUM-EWMA and mixed EWMA-CUSUM control charts for AR(1) process mean monitoring; and Osei-Aning et al. (2017b) provided the optimal design for the EWMA and CUSUM control charts for AR(1) processes.

Among other works considering AR(1) processes we have Chang and Wu (2011), which provided a Markov Chain approach to compute the ARL and Dasdemir et al. (2016) who evaluated the effect of outliers in the Phase II control charts performance. A recent contribution has been done by Weiß et al. (2018), which provide both parametric and non-parametric bootstrap methods to adjust control chart limits to have a guaranteed conditional in-control performance for AR(1) processes when parameters are unknown. The practical problem of building a time series model and that of the effect of model accuracy on the performance of control charts is discussed in Ledolter and Bisgaard (2011) and Zhou and Goh (2016), respectively.

It is well known that control charts when parameters are estimated underperform, as pointed out in Jensen et al. (2006) and Psarakis et al. (2014) literature reviews. Most of the works related to the monitoring of the process variance when parameters are estimated has been done under the assumption of independent observations. For example, Chen (1998), Maravelakis et al. (2002), Castagliola et al. (2009), Huwang et al. (2009), Maravelakis and Castagliola (2009), Castagliola and Maravelakis (2011), Zwetsloot et al. (2015), to mention a few. Recently, there was an increase on the number of researches concerning: (i) the effect of parameter estimation on control chart performance by means of the $AARL$ and $SDARL$, performance measures introduced by Jones and Steiner (2012), in order to take into account the Phase I sampling variability, also called the "practitioner-to-practitioner variability"; and, (ii) the guaranteed conditional in-control performance, for which Gandy and Kvaløy (2013) provided a bootstrapping method. Since then, several works were devoted to these research topics. Among the works concerning variance monitoring are: Epprecht et al. (2015), Faraz et al. (2015), Goedhart et al. (2017a), Guo and Wang (2017) and Faraz et al. (2018). A recent contribution that seems to be of strong practical impact is that of Aparisi et al. (2018), which are the first ones to design a control chart guaranteeing both the in-control and out-of-control performances.

It can be seen that the study of the performance of control charts for the *variance* monitoring of autocorrelated processes has not been deeply studied, although there are several studies about SPC techniques for autocorrelated process, and the effects of parameter estimation alongside the guaranteed conditional in-control performance in the independent data case. In this paper, the performance of the modified S^2 control chart when the process variance is estimated is evaluated. The rest of the chapter is organized as follows: Section 5.2 stands for the previous and proposed approaches to evaluate the performance of the modified S^2 control chart for AR(1) processes. The simulation study carried out to evaluate the conditional in-control performance is contained in Section 5.3 followed by the adjustment of the control limits in Section 5.4, where the effect on the out-of-control ARL is also studied. A numerical example is shown in Section 5.5 whereas conclusions and future works are discussed in Section 5.6.

5.2 The modified S^2 control chart for AR(1) processes

First of all, a brief description of the S^2 control chart for independent and identically distributed, i.i.d. normal observations is presented. Under that scenario, we have that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, and therefore, the one-sided S^2 control chart is set up by considering an upper control limit of the form: $UCL = \frac{\sigma^2}{n-1}L$, where L is a constant chosen accordingly to have a desired ARL_0. If α denote the probability of a type I error, then the one-sided S^2 control chart for the normal i.i.d. case signals whenever $S^2_i > \frac{\sigma^2}{n-1}\chi^2_{1-\alpha,n-1}$, where $\chi^2_{1-\alpha,n-1}$ is the $(1-\alpha)$-th quantile of a χ^2_{n-1} random variable. Unfortunately, for autocorrelated processes the exact distribution of S^2 (or of some transformation of it) is unknown, and other techniques must be applied in that case, as it is detailed in the following subsection.

5.2.1 Previous approach: known variance case

Amin et al. (1997) followed the suggestion of Vasilopoulos and Stamboulis (1978) of modifying the control limits by considering the process true variance. Given that the distribution of S^2 is not known, they did the modification by considering approximations to the probability density function, (p.d.f.) and the cumulative density function, (c.d.f.) of a function of S^2. In order to do this, they wrote the function $(n-1)S^2/\sigma^2$ as a quadratic form and they used the approximations of its p.d.f. and c.d.f. found in Mathai and Provost (1992).

Amin et al. (1997) they considered the common and widely used AR(1) model, given by

$$X_{i,j} - \mu_0 = \phi\left(X_{i,j-1} - \mu_0\right) + \varepsilon_{i,j}, \tag{5.2.1}$$

where $\varepsilon_{i,j} \sim N\left(0, \sigma_\varepsilon\right)$ are uncorrelated normal random variables; $\phi \in (-1,1)$ to ensure that the process is stationary; and $X_{i,j}$ and $X_{i+1,j}$ are uncorrelated, $\forall j = 1,2,\ldots,n$, that is to say, the autocorrelation is present within but not between samples. This sampling scheme is often applied in practice: independence between samples is suitable to be assumed when they are collected over sufficiently separated intervals of time. Under this model, there is a relationship between the process and errors variances given by:

$$\sigma_0^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}.$$

Amin et al. (1997) rewrote the chart statistic as:

$$\frac{(n-1)S_i^2}{\sigma_0^2} = \frac{(n-1)S_i^2}{\frac{\sigma_\varepsilon^2}{1-\phi^2}} = \frac{\sum_{j=1}^{n}\left(x_{i,j} - \bar{x}_i\right)^2}{\frac{\sigma_\varepsilon^2}{1-\phi^2}} = \sum_{j=1}^{n}\left(\frac{x_{i,j}}{\sqrt{\frac{\sigma_\varepsilon^2}{1-\phi^2}}} - \frac{\bar{x}_i}{\sqrt{\frac{\sigma_\varepsilon^2}{1-\phi^2}}}\right)^2 = \sum_{j=1}^{n}\left(y_{i,j} - \bar{y}_i\right)^2$$

where $y_{i,j} = x_{i,j}/\sqrt{\frac{\sigma_\varepsilon^2}{1-\phi^2}}$, and wrote this equation in a quadratic form:

$$\frac{(n-1)S_i^2}{\sigma_0^2} = Y_i'\left(I - \frac{1}{n}J\right)Y_i = Y_i'AY_i \tag{5.2.2}$$

where: $Y_i' = \left(y_{i,1}, y_{i,2}, \ldots, y_{i,n}\right)'$, I is the identity matrix of order n, J is an $n \times n$ matrix whose all entries are ones, and $A = I - \frac{1}{n}J$ is a positive semidefinite matrix of rank $(n-1)$. Given the model stated in equation (5.2.1), it follows that $X_i \sim N_n\left(\mathbf{0}, \sigma_0^2\boldsymbol{\Sigma}\right)$, the normal multivariate distribution, and therefore, $Y_i \sim N_n\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$. Now, the distribution of $Q(Y) = \frac{(n-1)S^2}{\sigma_0^2}$ could be approximated using the theorem found in Mathai and Provost (1992), which is stated in the Appendix and it is based on power series and generalized Laguerre polynomials.

Let β denote the conditional probability of a type II error, then it is given by

$$\beta = \mathbb{P}\left(S_i^2 \leq UCL|\hat{\sigma}_0^2\right) = \mathbb{P}\left(\frac{(n-1)S_i^2}{\sigma_0^2} \leq \frac{\hat{\sigma}_0^2}{\sigma_0^2}L\right) = F_{Q(Y)}\left(\frac{\hat{\sigma}_0^2}{\sigma_0^2}L\right).$$

As autocorrelation within samples and not between samples is assumed, then the Run Length, RL, follows a geometric distribution with parameter $1-\beta$, and, therefore, $ARL = 1/(1-\beta)$. With this result, the authors calculated the ARL for several combinations of Phase II sample sizes (n) and ϕ_0, finding the necessity of modifying the control limits since ARL_0 values lower than the nominal ARL_0 were obtained whenever $\phi_0 \neq 0$.

In order to modify the control limits to achieve a nominal ARL_0, the authors solved for L the following non-linear equation for several combinations of n and ϕ_0:

$$ARL_0 = \frac{1}{1-\beta} = \frac{1}{1 - F_{Q(Y)}\left(\frac{\hat{\sigma}_0^2}{\sigma_0^2} L\right)}, \tag{5.2.3}$$

but considering that $\widehat{\sigma}_0^2 = \sigma_0^2$. We will denote those modified control limits here as $L(n, \phi_0)$, since they depend only on n and ϕ_0. The calculation of the ARL via generalized Laguerre polynomials and the modification of the control limits are remarkable contributions from Amin et al.'s paper. However, the case of estimated parameters was not deeply studied and the modification (adjustment) of the control limits was not done in terms of guaranteeing a conditional in-control performance. These issues will be explored in Section 5.3 whereas the proposed model is detailed in the following subsection.

5.2.2 The proposed approach: estimated variance case

Considering the transformation of the chart statistic found in Amin et al. (1997), we are motivated to introduce an equivalent model. Consider that, at the i-th sampling point, the sequence of observations $\{X_{i,1}, X_{i,2}, \ldots, X_{i,n}\}$ follow the model:

$$\frac{X_{i,j} - \mu_0}{\tau\sigma_0} = \phi_0 \left(\frac{X_{i,j-1} - \mu_0}{\tau\sigma_0}\right) + \varepsilon_{i,j}, \tag{5.2.4}$$

for $j = 1, 2, \ldots, n$, where n stands for the subgroup size, μ_0 and σ_0 are the in-control process mean and standard deviation, respectively, whereas ϕ_0 is the in-control autoregressive parameter which lies on the interval $(-1, 1)$. The random variables $\varepsilon_{i,j} \sim N(0, \sigma_\varepsilon)$ are assumed to be uncorrelated, $\tau^2 = \sigma_1^2/\sigma_0^2$ stands for changes in variance and $X_{i,0}$ has the steady-state distribution, i.e. $X_{i,0} \sim N(\mu_0, \tau\sigma_0)$.

The same sampling scheme introduced before is considered for this model, that is to say, for $j = 1, 2, \ldots, n$, $X_{i,j}$ and $X_{i+1,j}$ are regarded as independent. Furthermore, as samples are assumed to be independent, the model in (5.2.4) could be rewritten (in order to have a simpler notation) by saying that the i-th sample follows the model:

$$\frac{X_j - \mu_0}{\tau\sigma_0} = \phi_0 \left(\frac{X_{j-1} - \mu_0}{\tau\sigma_0}\right) + \varepsilon_j. \tag{5.2.5}$$

It can be shown that under this model, the autoregressive parameter ϕ_0 is related to the errors variance by means of:

$$\sigma_\varepsilon^2 = 1 - \phi^2. \tag{5.2.6}$$

Even though this model has not been considered before, we think that it is suitable due to the changes in the process variance are not *defined* in terms of changes in the errors variance. However, as stated before, both models are equivalent and the first one could be considered in order to maintain the *status quo*.

The main objective is to detect departures from the in-control process variance, σ_0^2, and evaluate the capability detection of the modified S^2 control chart. Particularly, increases on the variance are wanted to be detected, so, the case $\tau^2 \leq 1$ is not considered as an out-of-control behavior and then, a one-sided modified S^2 control chart is considered here. In the next section, the methodology applied to evaluate the conditional in-control performance of the modified S^2 control chart is introduced as well as the adjustment of the control limits with the corresponding performance study.

5.3 Performance of the modified S^2 control chart when σ_0^2 is estimated

Since the distribution of the sample variance, S^2, is unknown for time series models, (just only a few moments of its distribution are known; Anderson, 2011), the approximations given in Mathai and Provost (1992) (stated here in the Appendix) will be considered. These approximations were previously validated by Amin et al. (1997) and they were again validated by the author of the manuscript.

There are two approaches to compute the $AARL$, $SDARL$ and other quantiles related to the ARL distribution: either to use its p.d.f. or to run extensive Monte Carlo simulations until a desired estimation accuracy is reached. The former approach has been considered by several authors (Saleh et al., 2015a; Saleh et al., 2015b; and Zwetsloot and Woodall, 2017, to mention a few) were the numerical integrals involved in the expectations were solved by Gaussian Quadrature methods; the latter approach is considered here, due to the lack of an *exact* distribution of S^2 or of some transformation of it.

Before considering the conditional in-control performance of the modified S^2 control chart, we will explore the relationship between the ARL and the variance estimation. In order to do that, Algorithm 5 was used by setting $n = 5$, $ARL_0 = 200$, $N = 100$ (the number of terms of the series expansion considered), $\phi_0 \in \{\pm(0.1, 0.5, 0.9)\}$, and without loss of generality, $\mu_0 = 0$, $\sigma_0 = 1$. It was considered that $\tau^2 = 1$ for the ARL_0 whereas $\tau^2 = 2$ for the ARL_1.

When ϕ_0 is fixed, it can be seen from Figures 5.3.1 to 5.3.4: (i) that the overestimation (underestimation) of σ_0^2 leads to higher (smaller) ARL values, something expected due to the widening (shrinking) of the UCL, and (ii) that the impact of overestimating the variance is stronger than that due to underestimating it, since the change on the ARL_0 is greater (less) than 1000 (200) when σ_0^2 is overestimated (underestimated) by 0.5 units, (see Figures 5.3.1 and 5.3.2). A similar behaviour is observed for the ARL_1. On the other hand, when $\hat{\sigma}_0^2$ is fixed, it can be inferred: (iii) that the ARL is not an increasing function of ϕ_0, in opposition with the findings for the $\bar{X}$ control chart. Moreover, (iv) that for positive or negative values of ϕ_0, the ARL curves (for different ϕ_0 values) intersect at $\widehat{\sigma}_0^2 = \tau^2$ were they change their order relation. For instance, in Figure 5.3.3, the ARL_1 for $\phi_0 = -0.9$ is greater than for $\phi_0 = -0.5$ which is greater than for $\phi_0 = -0.1$ whenever $\widehat{\sigma}_0^2 < \tau^2 = 2$, and the order is reversed when $\widehat{\sigma}_0^2 > \tau^2 = 2$. A similar behaviour is observed for positive values of ϕ_0 and for the ARL_0.

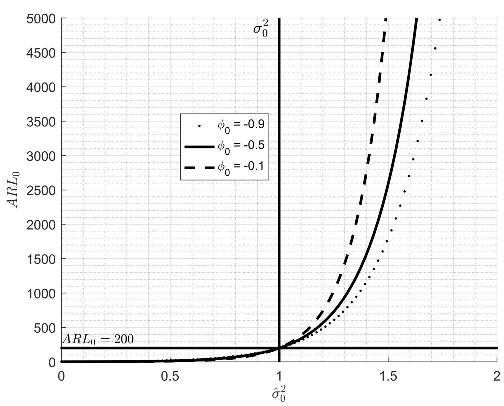


Figure 5.3.1: ARL_0 vs $\widehat{\sigma}_0^2$, for $n = 5, \tau^2 = 1$ and $\phi < 0$, (in-control process)

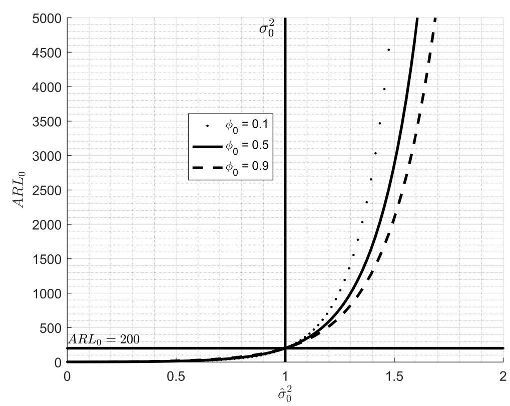


Figure 5.3.2: ARL_0 vs $\widehat{\sigma}_0^2$, for $n = 5, \tau^2 = 1$ and $\phi > 0$, (in-control process)

When implementing the control chart, each practitioner will collect a different Phase I sample, therefore, they will have a different variance estimation, leading to a different control limit, and then, to a different ARL. In order to consider the conditional in-control performance of the modified S^2 control chart, Algorithm 5 was used by setting $m \in \{25, 50, 100, 500\}$, $n = 5$, $ARL_0 = 200$, $\mu_0 = 0$, $\sigma_0 = 1$, $\tau^2 = 1$, $rep = 1000$, $N = 100$, $\phi_0 \in \{\pm(0.1, 0.5, 0.9)\}$.

From Table 5.3.1 the following conclusions (which are true in almost all cases) arises

- $AARL$ and $SDARL$ tend to decrease as m increases, for fixed ϕ_0.
- The $AARL$ and $SDARL$ tends to be higher as $|\phi_0| \to 1$.
- The median of the ARL_0, $MARL_0$, is below the nominal $ARL_0 = 200$.

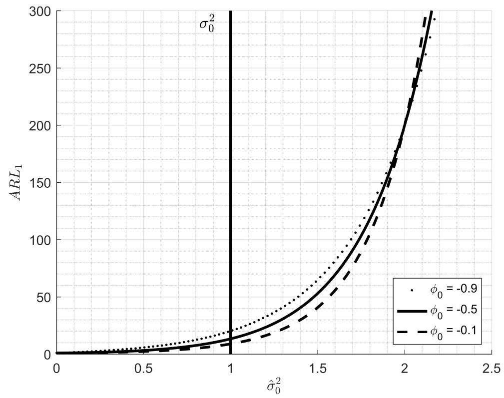


Figure 5.3.3: ARL_1 vs $\widehat{\sigma}_0^2$, for $n = 5, \tau^2 = 2$ and $\phi < 0$, (out-of-control process)

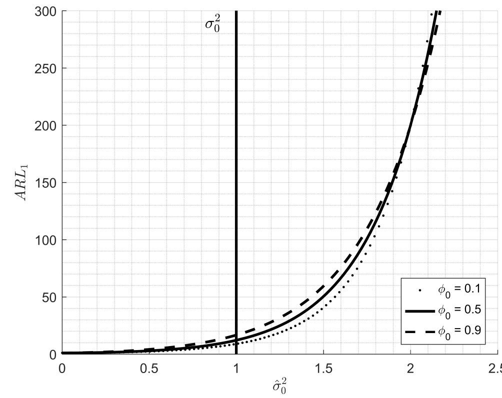


Figure 5.3.4: ARL_1 vs $\widehat{\sigma}_0^2$, for $n = 5, \tau^2 = 2$ and $\phi > 0$, (out-of-control process)

Algorithm 5 Calculation of $AARL$ and $SDARL$

Define m, n, μ_0, σ_0, ϕ_0, ARL_0, τ^2, rep, N.
$L \leftarrow L(n, \phi_0)$.
for $r = 1$ **to** rep **do**
Generate $X_1, X_2, \ldots, X_m$ based on the model

$$\frac{X_t - \mu_0}{\tau\sigma_0} = \phi_0 \left(\frac{X_{t-1} - \mu_0}{\tau\sigma_0} \right) + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, \sigma_\varepsilon),\ t = 1, \ldots, m.$$

$\widehat{\sigma}_{0,r}^2 \leftarrow S_r^2$, the sample variance.

$$y_r \leftarrow \frac{\widehat{\sigma}_{0,r}^2}{\sigma_0^2} L$$

$$F_{Q(Y),r}(y_r) \leftarrow \int_0^{y_r} \frac{\left(\frac{x}{2\beta}\right)^{n/2-1} e^{-\frac{x}{2\beta}}}{2\beta\Gamma(n/2)} dx + \sum_{k=1}^{N} c_k \frac{(k-1)!}{\Gamma(n/2+k)} \left(\frac{y_r}{2\beta}\right)^{n/2} e^{-\frac{y_r}{2\beta}} L_k^{(n/2)}\left(\frac{y_r}{2\beta}\right).$$

$$ARL_r \leftarrow \frac{1}{1 - F_{Q(Y),r}(y_r)}.$$

end for
$AARL \leftarrow \text{mean}(ARL_1, \ldots, ARL_{rep})$.
$SDARL \leftarrow \text{stdev}(ARL_1, \ldots, ARL_{rep})$.

- $AARL_0 > MARL_0$, implying that the ARL_0 distribution might be right-skewed.

Even though we consider the modified control limits reported by Amin et al. (1997), it can be seen that the $MARL_0$ is below the nominal ARL_0 in almost all cases studied here. This means that more than 50% of the times the control charts used by practitioners will show a performance below the expected one and that the adjustment made by the authors is not enough to jointly overcome the issue of autocorrelation and estimated parameters.

In the next section, the general guidelines to apply the bootstrapping methodology introduced by Gandy and Kvaløy (2013) to adjust the control limits of the modified S^2 control chart to have a guaranteed conditional in-control performance are considered.

5.4 Guaranteed conditional in-control performance

Aside the work of Aparisi et al. (2018), almost all the papers prior this work were devoted to guarantee a conditional *in-control* performance, usually done by exact methods or bootstrap-based techniques. The

Table 5.3.1: Descriptive statistics for the ARL_0 distribution considering $n = 5$ and a nominal $ARL_0 = 200$ for different values of m and ϕ_0.

	m							
	25		50		100		500	
ϕ_0	$AARL_0$ ($SDARL_0$)	$MARL_0$	$AARL_0$ ($SDARL_0$)	$MARL_0$	$AARL_0$ ($SDARL_0$)	$MARL_0$	$AARL_0$ ($SDARL_0$)	$MARL_0$
-0.9	1.08×10^9 (2.41×10^{10})	82.68	77008.79 (921209.17)	110.15	7725.24 (101705.17)	152.19	328.92 (461.94)	192.19
-0.5	6000.86 (72978.41)	159.04	692.54 (3012.84)	195.73	341.87 (587.22)	191.40	224.60 (108.07)	197.87
-0.1	17878.63 (482352.34)	146.12	594.81 (1986.78)	183.11	342.14 (616.49)	188.87	222.20 (101.22)	200.85
0.1	4172.68 (59409.43)	169.52	608.39 (1926.99)	173.84	324.21 (487.17)	195.29	217.41 (99.37)	196.86
0.5	1242.73 (12673.31)	96.82	2258.46 (51639.08)	136.45	306.78 (493.25)	166.22	212.06 (102.62)	189.51
0.9	1102.56 (19711.30)	11.60	2280.66 (33439.42)	28.06	935.28 (5885.15)	63.68	256.51 (333.29)	150.76

success when developing an exact method often relies on the knowledge of the *exact* distribution of the chart statistic or a transformation of it (as in Goedhart et al., 2017b; Goedhart et al., 2017a; and Faraz et al., 2018), which is not our case. On the other hand, the Gandy and Kvaløy's methodology has been widely applied, (Hu and Castagliola, 2017, Saleh et al., 2015a, Aly et al., 2015b, Faraz et al., 2015).

The bootstrap methodology has been explained in previous chapters, but here it is stated again, for the sake of completeness. Let P denote the process model and $\boldsymbol{\theta}$ the vector of parameters of the model P. With this notation, $\widehat{\boldsymbol{\theta}}$ and $\hat{P}$ denotes the estimated model parameters and the estimated model, respectively. In addition, let $L\left(P, \boldsymbol{\theta}\right)$ denote the L value used to achieve a desired performance when data is generated from model P and the control limit is set using $\boldsymbol{\theta}$. $L\left(\hat{P}_b^*, \widehat{\boldsymbol{\theta}}_b^*\right)$ and $L\left(\hat{P}, \widehat{\boldsymbol{\theta}}_b^*\right)$ are defined analogously. Then, the general outline of the bootstrap method is:

1. From historical data $X_1, \ldots, X_m$, obtain $\widehat{\boldsymbol{\theta}}$ and $\hat{P}$.
2. Generate a bootstrap sample $X_1^*, \ldots, X_m^*$, from $\hat{P}$, and compute $\hat{P}^*$ and $\widehat{\boldsymbol{\theta}}^*$. Repeat B times to have $\hat{P}_1^*, \ldots, \hat{P}_B^*$ and $\widehat{\boldsymbol{\theta}}_1^*, \ldots, \widehat{\boldsymbol{\theta}}_B^*$.
3. Consider the bootstrap distribution of $L\left(\hat{P}_b^*, \widehat{\boldsymbol{\theta}}_b^*\right) - L\left(\hat{P}, \widehat{\boldsymbol{\theta}}_b^*\right)$ and denote by L_{α^*} its α^*-th quantile. Then, take the adjusted control limit as the quantity $L\left(\hat{P}, \widehat{\boldsymbol{\theta}}\right) - L_{\alpha^*}$

Note that $L\left(P, \boldsymbol{\theta}\right) = L\left(\hat{P}_b^*, \widehat{\boldsymbol{\theta}}_b^*\right) = L\left(n, \phi_0\right)$ because the process variance coincides with the variance estimation, and that was previously done in Amin et al. (1997). With that in mind, the adjusted L value is calculated as the α^*-th quantile of the bootstrap distribution of $L\left(\hat{P}, \widehat{\boldsymbol{\theta}}_b^*\right)$.

Despite the proven effectiveness of the bootstrapping methodology, this approach might be computationally heavy and that is why exact methods are preferred over bootstrap based methods. In order to reduce the computational time, the adjusted L-values were calculated using Algorithm 6. The rationale of calculating them in that way relies on the fact that Amin et al. have been previously solved the equation (5.2.3) when $\widehat{\sigma}_0^2 = \sigma_0^2$. Now, as the quotient $D = \widehat{\sigma}_0^2/\sigma_0^2$ might be regarded as the coefficient of the control limit L, trying to solve equation (5.2.3) for $D \cdot L$ will have a solution of the form $\frac{1}{D}L\left(n, \phi_0\right)$, where $L\left(n, \phi_0\right)$ is the adjusted control limit found by Amin et al. (1997). It is worthy to mention that this approach remarkably reduces the computational time as the non-linear equation is no longer required to be solved in each iteration. However, as different Phase I samples will lead to different variance estimation, the bootstrap algorithm was repeated several times. Algorithm 6 was used to compute the adjusted L-value by setting $n = 5$, $\mu_0 = 0$, $\sigma_0^2 = 1$, $ARL_0 = 200$, $\alpha^* = 0.9$, $B = 1000$ and $rep = 1000$, for different values of $m \in \{25, 50, 100, 500\}$ and $\phi_0 \in \{\pm\left(0.1, 0.5, 0.9\right)\}$.

Algorithm 6 Calculation of the adjusted L-values

Define m, n, μ_0, σ_0, ϕ_0, ARL_0, α^*, B, rep.
$\tau^2 \leftarrow 1$.
$\alpha \leftarrow \frac{1}{ARL_0}$.
$r \leftarrow 1$.
for $r = 1$ **to** rep **do**
  Generate $X_1, X_2, \dots, X_m$ based on the model:
$$\frac{X_j - \mu_0}{\tau\sigma_0} = \phi_0 \left(\frac{X_{j-1} - \mu_0}{\tau\sigma_0}\right) + \varepsilon_j, \text{ where } \varepsilon_j \sim N(0, \sigma_\varepsilon).$$
  $\widehat{\sigma}^2_{0,r} \leftarrow S^2_r$
  for $b = 1$ **to** B **do**
    Generate $X^*_1, X^*_2, \dots, X^*_m$ based on the model:
$$\frac{X_j - \mu_0}{\tau\widehat{\sigma}_{0,r}} = \phi_0 \left(\frac{X_{j-1} - \mu_0}{\tau\widehat{\sigma}_{0,r}}\right) + \varepsilon_j, \text{ where } \varepsilon_j \sim N(0, \sigma_\varepsilon).$$
    $\widehat{\sigma}^{2*}_{0,b} \leftarrow S^{2*}_b$
    $L(r, b) \leftarrow \frac{\widehat{\sigma}^2_{0,r}}{\widehat{\sigma}^{2*}_{0,b}} L(n, \phi_0)$
  end for
  $L_r \leftarrow$ is the α^*-th quantile of $L(r, b)$.
end for
$L \leftarrow \text{mean}(L_1, \dots, L_{rep})$.

Table 5.4.1 shows the averaged adjusted L values as well as $ARL_{0,0.1}$, the 10-th quantile of the simulated ARL_0, in order to see if the desired in-control performance is reached. It can be seen that in almost all cases it was achieved, and that the adjusted L value decreases as ϕ_0 increases. It is well known that there is a trade off between the in-control and out-of-control performances due to the widening of the control limits. In order to explore that effect, Algorithm 5 was run considering both the unadjusted (U) and the adjusted (A) control limits, calculated using the L values provided in Amin et al. (1997) and in Table 5.4.1, respectively. As the overestimation of σ_0^2 leads to higher ARL_1 values (see Figures 5.3.3 and 5.3.4), the $MARL_1$ and the $ARL_{1,0.9}$, the median and the 90-th quantile of the conditional ARL_1 distribution, respectively, are reported in Table 5.4.2 to indicate that 50% and 90%, respectively of the control charts will have an ARL_1 of *at most* that value. Finally, the last column contains the ARL_1 values corresponding to the known variance case. It can be seen that the effect on the ARL_1 is reduced as m is increased, but even

Table 5.4.1: The averaged adjusted L values for the control limits for different values of m and ϕ_0 for a desired $ARL_0 = 200$ and $n = 5$. $ARL_{0,0.1}$ is the 10-th quantile of the simulated ARL_0 obtained using these values.

	m								
	25		50		100		500		
ϕ_0	L	$ARL_{0,0.1}$	L	$ARL_{0,0.1}$	L	$ARL_{0,0.1}$	L	$ARL_{0,0.1}$	$L(n, \phi_0)$
-0.9	104.80	194.60	77.08	201.51	60.64	205.13	43.19	209.00	33.17
-0.5	34.71	206.25	29.65	195.22	26.67	204.32	23.23	202.26	20.87
-0.1	23.74	186.57	20.66	207.03	18.89	203.15	16.91	197.61	15.54
0.1	22.25	188.24	19.21	203.70	17.50	217.24	15.60	205.28	14.33
0.5	22.05	176.06	17.90	196.68	15.67	195.44	13.37	204.41	11.95
0.9	19.88	222.01	12.26	219.91	8.49	205.80	5.30	195.54	3.92

more than $m = 500$ observations might be needed in order to mitigate that effect. The feasibility of that amount of data depends on the kind of process that we are monitoring, but for traditional manufacturing processes it might be regarded as infeasible.

Table 5.4.2: The $MARL_1$ and $ARL_{1,0.9}$, the 90-th quantile of the simulated ARL_1 (below within parenthesis) values considering a shift of $\tau^2 = 2$, with $n = 5$, $ARL_0 = 200$.

	m								
	25		50		100		500		
ϕ_0	U	A	U	A	U	A	U	A	Known Case
-0.9	12.47 (208.12)	600.22 (7.85×10^6)	14.61 (129.74)	216.55 (14543.53)	17.38 (82.97)	106.72 (1332.99)	19.68 (38.64)	39.42 (92.90)	20.14
-0.5	11.78 (53.75)	68.11 (762.25)	13.18 (34.52)	41.38 (156.33)	13.02 (25.75)	28.30 (68.24)	13.26 (18.10)	18.31 (26.09)	13.33
-0.1	7.71 (29.84)	39.92 (355.17)	8.54 (19.70)	23.65 (75.74)	8.66 (15.29)	16.17 (33.51)	8.90 (11.49)	11.47 (15.00)	8.89
0.1	8.23 (30.17)	42.05 (292.54)	8.33 (20.78)	22.98 (69.67)	8.78 (15.22)	16.65 (33.17)	8.81 (11.48)	11.32 (14.74)	8.87
0.5	8.25 (30.00)	64.02 (819.98)	9.89 (28.83)	39.44 (176.11)	10.98 (22.54)	27.54 (67.37)	11.78 (16.23)	16.52 (24.07)	12.12
0.9	3.32 (17.63)	319.89 (186575.63)	5.51 (26.85)	190.15 (20615.25)	8.79 (37.69)	87.10 (1100.35)	14.24 (28.65)	32.99 (81.39)	16.64

5.5 Numerical Example

A numerical example using simulated data is shown here, in order to detail the methodology used here. First of all, $m = 100$ subsequent observations following the model in equation (5.2.5) with $\mu_0 = 0$, $\sigma_0 = 1$, $\tau^2 = 1$ and $\phi_0 = 0.5$ were generated. This data is shown in Table 5.5.1 and constitutes the Phase I sample used to estimate the process variance, which results in $\hat{\sigma}_0^2 = 0.9038$. After that, in order to check for process stability a Phase I modified S^2 control chart is designed, considering $L = 15.67$ (taken from Table 5.4.1), and $n = 5$ so that $\widehat{UCL}_A = \dfrac{\hat{\sigma}_0^2}{n-1} L = 3.5408$. If the control limit is not adjusted, then $L = 11.95$ (taken from Amin et al., 1997) and $\widehat{UCL}_U = 2.7002$. The chart statistic is the sample variance, where each sample consisting on 5 consecutive observations is collected every 15 units. The corresponding values used to compute the variances are in bold in Table 5.5.1 and the control chart is depicted in Figure 5.5.1a. With this, we could assume that the process is stable and the estimations might be used for a Phase II online monitoring. After

Table 5.5.1: Phase I data for the numerical example.

j											S^2
1-10	**0.08**	**-0.78**	**-0.03**	**0.57**	**1.03**	-0.08	0.35	0.26	0.85	0.89	**0.46**
11-20	1.22	0.50	0.12	0.93	-1.37	-1.12	-1.66	-1.16	-0.02	0.71	
21-30	**-0.53**	**-0.67**	**-0.22**	**-0.36**	0.08	0.39	-0.61	-0.46	-2.08	-0.05	**0.08**
31-40	-0.57	-1.33	-0.88	-1.68	-0.86	-0.91	1.43	1.70	-1.31	-0.27	
41-50	**-1.35**	**-0.89**	**-0.31**	**0.50**	**0.01**	1.37	0.27	0.42	0.78	0.47	**0.53**
51-60	1.00	0.78	-0.29	-1.71	0.76	-0.15	0.02	0.50	0.35	-0.61	
61-70	**-0.71**	**-0.46**	**1.05**	**-0.22**	**0.57**	0.55	0.07	-0.88	-0.69	-0.42	**0.55**
71-80	-1.48	-0.57	-1.00	-0.58	0.00	-0.78	-0.64	-0.02	-1.60	0.10	
81-90	**2.15**	**1.91**	**0.68**	**0.71**	**-0.54**	1.36	1.49	1.43	-0.04	0.25	**1.18**
91-100	-0.36	-0.45	-0.72	-1.25	-1.41	-0.89	-1.91	-0.43	-0.32	0.45	

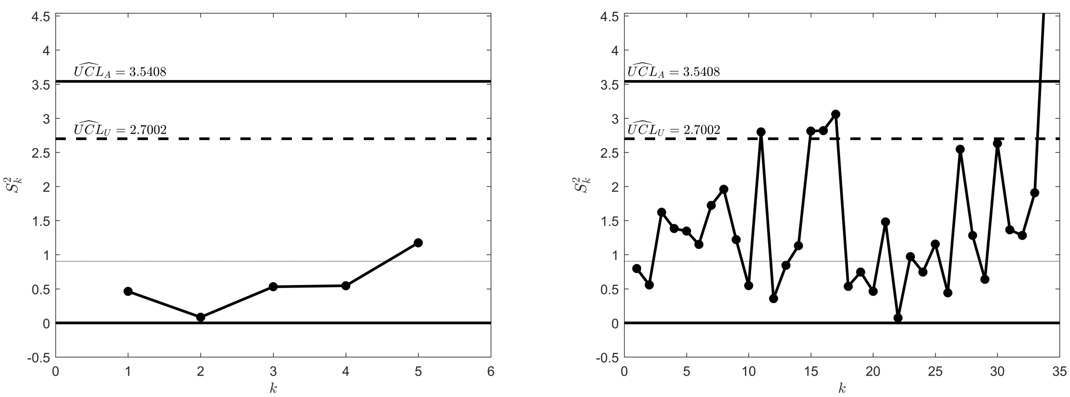
that, observations are generated according to the model (5.2.5) with $\mu_0 = 0$, $\sigma_0^2 = 1$, $\phi_0 = 0.5$, *but* $\tau^2 = 2$, corresponding to a shift in the process variance. The collected samples are presented in Table 5.5.2 with their respective variances in the last column. The effect of the adjustment of the control limit can be seen: for the unadjusted control limit case, the chart signals at samples # 11, 15, 16, 17 and 34, in opposition to the adjusted control limit case, where the first signal is triggered at the sample # 34.

5.6 Conclusions and future work

Amin et al. (1997) developed the modified S^2 control chart for monitoring the variance of AR(1) processes. Even though they provide a way to compute the ARL avoiding the use of simulation, they did not take into account the variance estimation on their study. In this research, a re-evaluation of the modified S^2 control chart was carried out, finding that even when considering the modified S^2 control chart, more that 50% of

Table 5.5.2: Phase II data for the numerical example.

k	X_1	X_2	X_3	X_4	X_5	S_k^2	k	X_1	X_2	X_3	X_4	X_5	S_k^2
1	-1.26	-0.94	-2.08	-2.09	0.06	**0.80**	18	0.63	0.75	0.10	-1.04	-0.31	**0.54**
2	1.07	-0.54	-0.08	-0.58	0.72	**0.56**	19	-0.53	-2.17	-0.46	-0.43	0.11	**0.75**
3	-1.09	1.36	-1.22	-0.25	-1.98	**1.62**	20	-0.74	0.47	-0.50	0.19	0.91	**0.46**
4	-1.03	-1.83	0.98	0.29	0.54	**1.38**	21	-0.18	-1.22	-0.32	1.16	-2.09	**1.48**
5	-2.04	-2.20	-0.67	-0.46	0.58	**1.35**	22	0.16	0.25	0.56	-0.19	0.32	**0.07**
6	1.64	-0.31	-1.34	-0.11	-0.20	**1.15**	23	-0.78	-0.08	1.37	0.57	-1.04	**0.97**
7	0.02	-1.08	-1.70	-0.30	1.76	**1.72**	24	-0.15	0.46	-0.10	1.98	0.45	**0.75**
8	-0.49	1.11	2.23	2.43	3.07	**1.96**	25	-0.60	1.70	1.21	1.26	2.26	**1.16**
9	0.51	-1.72	-1.87	-1.12	-2.34	**1.22**	26	-0.93	-0.69	0.72	-0.62	-0.02	**0.44**
10	-0.74	1.19	-0.36	0.07	0.38	**0.55**	27	-1.65	-0.20	2.64	1.12	0.07	**2.55**
11	2.61	1.00	1.29	-0.86	-1.51	**2.80**	28	2.35	0.03	-0.45	-0.26	0.61	**1.28**
12	0.46	0.87	-0.14	0.27	-0.69	**0.36**	29	-1.54	-1.96	-0.60	-0.53	-2.32	**0.64**
13	0.81	-0.45	0.08	0.47	1.99	**0.84**	30	-2.44	-3.34	-1.65	-0.27	0.68	**2.63**
14	-1.57	-0.52	-2.38	-1.03	0.44	**1.13**	31	-1.50	-3.21	-0.88	-0.54	-0.29	**1.37**
15	-1.51	-1.01	1.00	2.53	1.27	**2.81**	32	-0.57	-0.67	1.99	0.81	1.03	**1.28**
16	2.10	0.86	-0.64	-0.45	-2.36	**2.82**	33	-3.07	-3.36	-2.64	-1.14	-0.13	**1.91**
17	-1.17	-1.35	1.02	0.78	2.89	**3.06**	34	3.55	-0.38	-2.31	-1.50	-1.81	**5.60**



(a) Phase I modified S^2 control chart.

(b) Phase II modified S^2 control chart.

Figure 5.5.1: The modified S^2 control charts for simulated data, with unadjusted control limit (dashed line) and adjusted control limit (solid line).

the times the charts designed by practitioners will perform worse than expected due to the variance estimation.

Considering the practitioner-to-practitioner variability, a bootstrapping methodology based on Gandy and Kvaløy's method, was applied in order to guarantee, with a certain probability, a conditional in-control performance. Results shown that such performance is reached but the effect on the ARL_1 is quite strong. In fact, more than $m = 500$ Phase I observations are required to mitigate this effect.

The design and performance of a S^2 control chart when all parameters are estimated is left as a future work. It is possible to design a Shewhart-type control chart for S^2 when monitoring AR(1) processes, since the first two moments of the S^2 distribution are known (Anderson, 2011) and there is a methodology to compute the chart constant to achieve a desired ARL_0 with the approximation provided in Amin et al. (1997) and that was used here.

Appendix

In this section, the theorem used to approximate the distribution of S^2 when dealing with AR(1) processes is detailed. This theorem was found in Amin et al. (1997) and in Mathai and Provost (1992). Here, it is written keeping the notation used throughout the paper.

Theorem 1. *Let $f_{Q(Y)}(y)$ and $F_{Q(Y)}(y)$ be the p.d.f. and c.d.f. of the quadratic form $Q(Y) = Y'AY$, respectively, where A is a symmetric matrix positive definite and $Y \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ a positive definite matrix with entries given by $\phi^{|i-j|}$. It can be shown that:*

1. $f_{Q(Y)}(y) = \sum_{k=0}^{\infty} c_k \frac{k!}{2\beta\Gamma(n/2+k)} \left(\frac{y}{2\beta}\right)^{n/2-1} e^{-\frac{y}{2\beta}} L_k^{(n/2-1)}\left(\frac{y}{2\beta}\right)$, $y > 0$

2. $F_{Q(Y)}(y) = \int_0^y \frac{\left(\frac{x}{2\beta}\right)^{n/2-1} e^{-\frac{x}{2\beta}}}{2\beta\Gamma(n/2)} dx + \sum_{k=1}^{\infty} c_k \frac{(k-1)!}{\Gamma(n/2+k)} \left(\frac{y}{2\beta}\right)^{n/2} e^{-\frac{y}{2\beta}} L_k^{(n/2)}\left(\frac{y}{2\beta}\right)$, $y > 0$

where:

- $\lambda_j, j = 1, 2, \ldots, n$ *are the eigenvalues of* $\boldsymbol{\Sigma}^{1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{1/2}$.
- *The vector* $\boldsymbol{b} = \boldsymbol{P}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ *where* ***P*** *is an orthogonal matrix with* $\boldsymbol{P}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{P} = diag(\lambda_1, \ldots, \lambda_n)$
- $\beta = \frac{\lambda_{max} + \lambda_{min}}{2}$, *where* $\lambda_{max}(\lambda_{min})$ *is the largest (smallest) positive eigenvalue.*
- $c_0 = 1$*; the other coefficients are calculated recursively using* $c_k = \frac{1}{k}\sum_{r=0}^{k-1} d_{k-r}c_r$, *for* $k \geq 1$.
- $d_k = \frac{1}{2}\left[-\frac{k}{\beta}\sum_{j=1}^{n} \lambda_j b_j^2 \left(1 - \frac{\lambda_j}{\beta}\right)^{k-1} + \sum_{j=1}^{n}\left(1 - \frac{\lambda_j}{\beta}\right)^k\right]$, *for* $k \geq 1$.
- $L_k^{(\alpha)}(x) = \frac{1}{k!}e^x x^{-\alpha}\left[\frac{d^k}{dx^k}\left(e^{-x}x^{k+\alpha}\right)\right]$ *for* $\alpha > -1, k \in \mathbb{N}$, *denotes the k-th generalized (or associated) Laguerre polynomial.*

If A is positive semidefinite with $rank(A) = p < n$, substitute n by p.

For the case of $\boldsymbol{\mu} = \mathbf{0}$, some of the parameters are reduced. For instance:

- The vector **b** $= \mathbf{P}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} = \mathbf{0}$, meaning that $b_j = 0, \forall j$, and the computation of the matrix **P** is unnecessary.
- The coefficients d_k are calculated as $d_k = \frac{1}{2}\left[\sum_{j=1}^{n}\left(1 - \frac{\lambda_j}{\beta}\right)^k\right]$, for $k \geq 1$.

Note: In this research, the first N terms of the series expansion are considered, as suggested by Amin et al. (1997).

Chapter 6. General conclusions and future work

6.1 General conclusions

The effect of parameter estimation on the performance of control charts for autocorrelated data was carried out through this research, focused on the modified $\bar{X}$ and S^2 control charts for AR(1) processes. Concerning the research questions stated in Section 1.4 and the research hypotheses stated in Section 1.5, we found that:

- Q_1.- What is the conditional performance of $\bar{X}$ control chart for monitoring the mean of AR(1) processes when using autocorrelation estimators?

  In general, the overestimation of ϕ_0 provides higher ARL values for either the process being under control or not. Moreover, it is more likely to have larger ARL_0 values and small ARL_1 ones when $\phi_0 < 0$. Nevertheless, even in such cases, around 50% of the times control charts have a smaller ARL_0 than the expected one.

  As the overestimation of ϕ_0 affects the ARL of the $\bar{X}$ chart, it is recommended to use an estimator with the smaller bias and standard deviation. We recommend to use the least-squares estimator, $\hat{\phi}_{0,LS}$.

- Q_2.- What is the conditional performance of $\bar{X}$ control chart for monitoring the mean of AR(1) processes under parameter estimation?

  The $\bar{X}$ control chart performs worst for positive autocorrelation than for negative one, since there is a widening (shrinking) on the control limits for $\phi_0 > 0$ ($\phi_0 < 0$) due to the coefficient C_2, even in the *known* parameters case. Results shown that

  1. The overestimation of ϕ_0 and/or σ_0 leads to higher ARL values.
  2. The effect on the estimation of μ_0 is symmetrical around the current process mean $\mu_0 + \delta\sigma_0$ and leads to lower ARL values despite it is over or underestimated.
  3. More than 50% of the times the ARL_0 of the control chart is below the nominal value.

  Concerning the estimators considered in this study, $\hat{\sigma}_{0,SQ}$ and $\hat{\phi}_{0,LS}$ are suggested to be used.

- Q_3.- What is the conditional performance of the modified S^2 control chart for monitoring the variance of AR(1) processes when the variance is estimated?

  Results show that the overestimation (underestimation) of σ_0^2 leads to higher (smaller) ARL values, and that the effect on the ARL is stronger for the overestimation than for the underestimation of the process variance.

  Even though the modified S^2 control chart was considered, more that 50% of the times the charts designed by practitioners will perform worse than expected due to the variance estimation.

- Q_4.- What is the effect on the performance (in terms of $AARL$ and $SDARL$) of $\bar{X}$ and S^2 control charts when applying the Gandy and Kvaløy's bootstrap methodology to adjust control limits to have a guaranteed conditional in-control performance?

  In general, it is possible to apply the Gandy and Kvaløy (2013) bootstrap methodology to guarantee a conditional in-control performance. The conditional in-control performance is ensured, but there are two side-effects: there is an increase on the $AARL$ and the $SDARL$ in both in-control and out-of-control performances.

  This technique might be computationally heavy, but some simplifications could be made, as the one made in Section 5.4.

6.2 Future work

There are several lines to follow, as was mentioned in Section 2.5 and in Chapters 3,4 and 5. Considering the work developed and applied here, we could say it is left as a future work:

- The extension of this work to another time series models such as AR(p) process with $p > 1$, MA processes and ARMA processes. (Nevertheless, this is not something trivial and easy to do, as was stated in Section 4.6, due to the distribution of $\bar{X}$ and S^2 are not easy to be obtained for processes different from the AR(1) processes and the required effort might not be justified since it seems that a wide variety of processes might be modeled as AR(1) processes).
- The extension of this work to CUSUM and EWMA control charts.
- Once the previous point is assessed, a head-to-head comparison between the Shewhart, CUSUM and EWMA control charts for AR(1) processes might be done in order to identify which control chart is preferred when monitoring AR(1) processes under estimated parameters.
- Concerning the variance monitoring, it is possible to design a Shewhart-type control chart for monitoring of AR(1) process, since the first two moments of the S^2 distribution are known and there is a way to compute values of its p.d.f. and c.d.f.
- If the control chart in the previous point is designed, then a joint $\bar{X}$-S^2 control chart for AR(1) processes under estimated parameters might be easily developed. However, the adjustment of the control limits would require more effort.
- Explore the effect of outliers on the conditional performance of the $\bar{X}$ control chart for the autocorrelation estimators considered here, where the robust ones might perform better than the least-squares based ones.

In addition to those works, another lines to follow include

- Explore the self-starting methodology applied to time series models (applied to the original autocorrelated observations or to the residuals of the fitted model).
- Head-to-head comparisons between Shewart, CUSUM and EWMA self-starting control charts for autocorrelated processes when the parameters are estimated (applied to the observations or to the residuals of the fitted model).
- Explore the feasibility of to develop exact methods to adjust control chart limits to have a guaranteed conditional in-control performance.
- Explore the feasibility of to develop or proposed a method to simultaneously guarantee the in-control and out-of-control performances.
- Explore the applicability of these methods (developed for AR(1) processes) to other ARIMA models.
- Explore the feasibility to design control charts for stationary processes without any assumption about the process model.

References

W. Albers and W.C.M. Kallenberg. New Corrections for Old Control Charts. *Quality Engineering*, 17(3): 467–473, 2005.

M.M. Ali. Distribution of the Least Squares Estimator in a First Order Autoregressive Model. *Econometric Review*, 21(1):89–119, 2002.

H. Alshraideh and E. Khatatbeh. A Gaussian Process Control Chart for Monitoring Autocorrelated Process Data. *Journal of Quality Technology*, 46(4):317, 2014.

L.C. Alwan. Effects of Autocorrelation on Control Chart Performance. *Communications in Statistics-Theory and Methods*, 21(4):1025–1049, 1992.

L.C. Alwan and D. Radson. Time-Series Investigation of Subsample Mean Charts. *IIE Transactions*, 24(5): 66–80, 1992.

L.C. Alwan and H.V. Roberts. Time-Series Modeling for Statistical Process Control. *Journal of Business & Economic Statistics*, 6(1):87–95, 1988.

A.A. Aly, M.A. Mahmoud, and W.H. Woodall. A Comparison of the Performance of Phase II Simple Linear Profile Control Charts when Parameters are Estimated. *Communications in Statistics-Simulation and Computation*, 44(6):1432–1440, 2015a.

A.A. Aly, N.A. Saleh, M.A Mahmoud, and W.H Woodall. A Re-Evaluation of the Adaptive Exponentially Weighted Moving Average Control Chart when Parameters are Estimated. *Quality and Reliability Engineering International*, 31(8):1611–1622, 2015b. doi: 10.1002/qre.1695.

A.A. Aly, M.A. Mahmoud, and R. Hamed. The Performance of the Multivariate Adaptive Exponentially Weighted Moving Average Control Chart with Estimated Parameters. *Quality and Reliability Engineering International*, 32(3):957–967, 2016. doi: 10.1002/qre.1806.

R.W. Amin, W. Schmid, and O. Frank. The Effects of Autocorrelation on the R-Chart and the s^2-Chart. *The Indian Journal of Statistics, Series B (1960-2002)*, 59(2):229–255, 1997.

T.W. Anderson. *The Statistical Analysis of Time Series*, volume 19. John Wiley & Sons, 2011.

F. Aparisi, J. Mosquera, and E.K. Epprecht. Simultaneously Guaranteeing the In-Control and Out-of-Control Performances of the S^2 Control Chart with Estimated Variance. *Quality and Reliability Engineering International*, 2018.

M. Bagshaw and R.A. Johnson. The Effect of Serial Correlation on the Performance of Cusum Tests II. *Technometrics*, 17(1):73–80, 1975.

D.P. Bischak and D. Trietsch. The Rate of False Signals in $\bar{X}$ Control Charts with Estimated Limits. *Journal of Quality Technology*, 39(1):54–65, 2007.

G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2015.

W.J. Braun. Run Length Distributions for Estimated Attributes Charts. *Metrika*, 50(2):121–129, 1999.

D. Brook and D.A. Evans. An Approach to the Probability Distribution of Cusum Run Length. *Biometrika*, 59(3):539–549, 1972.

G. Capizzi and G. Masarotto. An Adaptive Exponentially Weighted Moving Average Control Chart. *Technometrics*, 45(3):199–207, 2003.

G. Capizzi and G. Masarotto. The Ewmast Control Charts with Estimated Limits: Properties and Recommendations. In *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, pages 1403–1407. IEEE, 2007.

G. Capizzi and G. Masarotto. Adaptive Generalized Likelihood Ratio Control Charts for Detecting Unknown Patterned Mean Shifts. *Journal of Quality Technology*, 44(4):281, 2012.

P. Castagliola and P.E. Maravelakis. A CUSUM Control Chart for Monitoring the Variance when Parameters are Estimated. *Journal of Statistical Planning and Inference*, 141(4):1463–1478, 2011.

P. Castagliola and S. Wu. Design of the c and np Charts when the Parameters are Estimated. *International Journal of Reliability, Quality and Safety Engineering*, 19(2):1–16, 2012.

P. Castagliola, G. Celano, and G. Chen. The Exact Run Length Distribution and Design of the S^2 Chart when the In-Control Variance is Estimated. *International Journal of Reliability, Quality and Safety Engineering*, 16(1):23–38, 2009.

P. Castagliola, S. Wu, M.B.C. Khoo, and S. Chakraborti. Synthetic Phase II Shewhart-Type Attributes Control Charts when Process Parameters are Estimated. *Quality and Reliability Engineering International*, 30(3):315–335, 2014.

S. Chakraborti. Parameter Estimation and Design Considerations in Prospective Applications of the $\bar{X}$ Chart. *Journal of Applied Statistics*, 33(4):439–459, 2006.

S. Chakraborti and S.W. Human. Parameter Estimation and Performance of the p-Chart for Attributes Data. *IEEE Transactions on Reliability*, 55(3):559–566, 2006.

S. Chakraborti, S.W. Human, and M.A. Graham. Phase I Statistical Process Control Charts: An Overview and Some Results. *Quality Engineering*, 21(1):52–62, 2009.

Y.M. Chang and T.L. Wu. On Average Run Lengths of Control Charts for Autocorrelated Processes. *Methodology and Computing in Applied Probability*, 13(2):419–431, 2011.

G. Chen. The Run Length Distributions of the r, s and S^2 Control Charts when σ is Estimated. *Canadian Journal of Statistics*, 26(2):311–322, 1998.

Y. Cheng, L. Sun, and B. Guo. Phase ii Synthetic Exponential Charts and Effect of Parameter Estimation. *Quality Technology & Quantitative Management*, 15(1):125–142, 2018.

C.H. Chin and D.W. Apley. Performance and Robustness of Control Charting Methods for Autocorrelated Data. *Journal of the Korean Institute of Industrial Engineers*, 34(2):122–139, 2008.

A.F.B. Costa and P. Castagliola. Effect of Measurement Error and Autocorrelation on the $\bar{X}$ Chart. *Journal of Applied Statistics*, 38(4):661–673, 2011.

C. Croux, S. Gelper, and K. Mahieu. Robust Control Charts for Time Series Data. *Expert Systems with Applications*, 38(11):13810–13815, 2011.

S.V. Crowder. A Simple Method for Studying Run–Length Distributions of Exponentially Weighted Moving Average Charts. *Technometrics*, 29(4):401–407, 1987.

E. Dasdemir, C. Weiß, M.C. Testik, and S. Knoth. Evaluation of Phase I Analysis Scenarios on Phase II Performance of Control Charts for Autocorrelated Observations. *Quality Engineering*, 28(3):293–304, 2016.

A.B.A. Dawod, M. Riaz, and S.A. Abbasi. On Model Selection for Autocorrelated Processes in Statistical Process Control. *Quality and Reliability Engineering International*, 33(4):867–882, 2017.

B. De Ketelaere, T. Rato, E. Schmitt, and M. Hubert. Statistical Process Monitoring of Time-Dependent Data. *Quality Engineering*, 28(1):127–142, 2016.

M.D. Diko, R. Goedhart, S. Chakraborti, R.J.M.M. Does, and E.K. Epprecht. Phase ii Control Charts for Monitoring Dispersion when Parameters are Estimated. *Quality Engineering*, 29(4):605–622, 2017.

E.K. Epprecht, L.D. Loureiro, and S. Chakraborti. Effect of the Amount of Phase I Data on the Phase II Performance of S^2 and S Control Charts. *Journal of Quality Technology*, 47(2):139–155, 2015.

A. Faraz, W.H. Woodall, and C. Heuchenne. Guaranteed Conditional Performance of the S^2 Control Chart with Estimated Parameters. *International Journal of Production Research*, 53(14):4405–4413, 2015.

A. Faraz, C. Heuchenne, and E. Saniga. The np Chart with Guaranteed In-control Average Run Lengths. *Quality and Reliability Engineering International*, 33(5):1057–1066, 2017.

A. Faraz, C. Heuchenne, and E. Saniga. An Exact Method for Designing Shewhart $\bar{X}$ and s^2 Control Charts to Guarantee In-Control Performance. *International Journal of Production Research*, 56(7):2570–2584, 2018.

B.C. Franco, P. Castagliola, G. Celano, and A.F.B. Costa. A New Sampling Strategy to Reduce the Effect of Autocorrelation on a Control Chart. *Journal of Applied Statistics*, 41(7):1408–1421, 2014.

B.C. Franco, G. Celano, P. Castagliola, A.F.B. Costa, and M.A.G. Machado. A New Sampling Strategy for the Shewhart Control Chart Monitoring a Process with Wandering Mean. *International Journal of Production Research*, 53(14):4231–4248, 2015.

W.A. Fuller. *Introduction to Statistical Time Series*. John Wiley & Sons, 2nd edition, 1996.

A. Gandy and J.T. Kvaløy. Guaranteed Conditional Performance of Control Charts via Bootstrap Methods. *Scandinavian Journal of Statistics*, 40(4):647–668, 2013.

J.A. Garza-Venegas, V.G. Tercero-Gómez, A. Cordero-Franco, M. Temblador-Pérez, and M. Beruvides. An Evaluation of Change-point Estimators for a Sequence of Normal Observations with Unknown Parameters. *Communications in Statistics-Simulation and Computation*, 46(6):4297–4317, 2017. doi: 10.1080/03610918.2015.1115524.

J.A. Garza-Venegas, V.G. Tercero-Gómez, L. Lee Ho, P. Castagliola, and G. Celano. Effect of Autocorrelation Estimators on the Performance of the $\bar{X}$ Control Chart. *Journal of Statistical Computation and Simulation*, 88(13):2612–2630, 2018. doi: 10.1080/00949655.2018.1479752.

B.K. Ghosh, M.R. Reynolds Jr., and V.H. Yer. Shewhart X-Charts with Estimated Process Variance. *Communications in Statistics-Theory and Methods*, 10(18):1797–1822, 1981.

R. Goedhart, da Silva M.M., M. Schoonhoven, E.K. Epprecht, S. Chakraborti, R.J.M.M. Does, and A. Veiga. Shewhart control charts for dispersion adjusted for parameter estimation. *IISE Transactions*, 49(8):838–848, 2017a.

R. Goedhart, M. Schoonhoven, and R.J.M.M. Does. Guaranteed in-control Performance for the Shewhart $\bar{X}$ and X Control Charts. *Journal of Quality Technology*, 49(2):155–171, 2017b.

M.A. Graham, A. Mukherjee, and S. Chakraborti. Distribution-Free Exponentially Weighted Moving Average Control Charts for Monitoring Unknown Location. *Computational Statistics and Data Analysis*, 56(8): 2539–2561, 2012.

B. Guo and B.X. Wang. The Design of the s^2 Control Charts Based on Conditional Performance via Exact Methods. *Quality and Reliability Engineering International*, 33(7):1567–1575, 2017.

J.N. Haddad. On Robust Estimation in the First Order Autoregressive Processes. *Communications in Statistics - Theory and Methods*, 29(1):45–54, 2000.

M. Hany and M.A. Mahmoud. An Evaluation of the Crosier's CUSUM Control Chart with Estimated Parameters. *Quality and Reliability Engineering International*, 32(5):1825–1835, 2016.

K. Harris, K. Triantafyllopoulos, E. Stillman, and T. McLeay. A Multivariate Control Chart for Autocorrelated Tool Wear Processes. *Quality and Reliability Engineering International*, 32(6):2093–2106, 2016.

T.J. Harris and W.H. Ross. Statistical Process Control Procedures for Correlated Observations. *The Canadian Journal of Chemical Engineering*, 69:48–57, 1991.

D.M. Hawkins. Self-starting Cusum Charts for Location and Scale. *The Statistician*, 36(4):299–316, 1987.

D.M. Hawkins and Q. Deng. A Nonparametric Change-Point Control Chart. *Journal of Quality Technology*, 42(2):165, 2010.

D.M. Hawkins and Q. Wu. The CUSUM and the EWMA head-to-head. *Quality Engineering*, 26(2):215–222, 2014.

D.M. Hawkins and K.D. Zamba. A Change-Point Model for a Shift in Variance. *Journal of Quality Technology*, 37(1):21–31, 2005.

D.M. Hawkins, P. Qiu, and C.W. Kang. The Changepoint Model for Statistical Process Control. *Journal of Quality Technology*, 35(4):355–366, 2003.

F. He, W. Jiang, and L. Shu. Improved Self-Starting Control Charts for Short Runs. *Quality Technology and Quantitative Management*, 5(3):289–308, 2008.

X. Hu and P. Castagliola. Guaranteed Conditional Design of the Median Chart with Estimated Parameters. *Quality and Reliability Engineering International*, 33(8):1874–1884, 2017.

L. Hurwicz. Least Squares Bias in Time Series. *Statistical inference in dynamic economic models*, 10: 365–383, 1950.

L. Huwang, Y.T. Wang, A.B. Yeh, and Z.J. Chen. On the Exponentially Weighted Moving Variance. *Naval Research Logistics (NRL)*, 56(7):659–668, 2009.

W.A. Jensen, L.A. Jones-Farmer, C.W. Champ, and W.H. Woodall. Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology*, 38(4):349–364, 2006.

D.R. Jeske. On the Conditional in-control ARL of a CUSUM Statistic. *Applied Stochastic Models in Business and Industry*, 32(5):558–564, 2016a.

D.R. Jeske. Determining the Reference Sample Size Needed to Control the Accuracy of the Conditional in-control ARL of a Normal-Theory CUSUM. *Quality and Reliability Engineering International*, 32(7): 2499–2504, 2016b.

W. Jiang, K.L. Tsui, and W.H. Woodall. A New Spc Monitoring Method: The Arma Chart. *Technometrics*, 42(4):399–410, 2000.

R.A. Johnson and M. Bagshaw. The Effect of Serial Correlation on the Performance of Cusum Tests. *Technometrics*, 16(1):103–112, 1974.

M.A. Jones and S.H. Steiner. Assessing the Effect of Estimation Error on Risk-Adjusted CUSUM Chart Performance. *International Journal for Quality in Health Care*, 24(2):176–181, 2012.

L.A. Jones-Farmer, C.W. Champ, and S.E. Rigdon. The Performance of Exponentially Weighted Moving Average Charts with Estimated Parameters. *Technometrics*, 43(2):156–167, 2001.

L.A. Jones-Farmer, C.W. Champ, and S.E. Rigdon. The Run Length Distribution of the CUSUM with Estimated Parameters. *Journal of Quality Technology*, 36(1):95–108, 2004.

L.A. Jones-Farmer, W.H. Woodall, S.H. Steiner, and C.W. Champ. An Overview of Phase I Analysis for Process Improvement and Monitoring. *Journal of Quality Technology*, 46(3):265, 2014.

M.J. Keefe, W.H. Woodall, and L.A. Jones-Farmer. The Conditional in-control Performance of Self-Starting Control Charts. *Quality Engineering*, 27(4):488–499, 2015.

M.G. Kendall. Note on Bias in the Estimation of Autocorrelation. *Biometrika*, 41(3-4):403–404, 1954.

S.H. Kim, C. Alexopoulos, K.L. Tsui, and J.R. Wilson. A Distribution-Free Tabular Cusum Chart for Autocorrelated Data. *IIE Transactions*, 39(3):317–330, 2007.

H. Kramer and W. Schmid. The Influence of Parameter Estimation on the ARL of Shewhart Type Charts for Time Series. *Statistical Papers*, 41(2):173–196, 2000.

J. Ledolter and S. Bisgaard. Challenges in Constructing Time Series Models from Process Data. *Quality and Reliability Engineering International*, 27(2):165–178, 2011.

B.H. Lee and W.W.S. Wei. The Use of Temporally Aggregated Data on Detecting a Mean Change of a Time Series Process. *Communications in Statistics-Theory and Methods*, 46(12):5851–5871, 2017.

H.C. Lee and D.W. Apley. Improved Design of Robust Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes. *Quality and Reliability Engineering International*, 27(3):337–352, 2011.

J. Lee, N. Wang, L. Xu, A. Schuh, and W.H. Woodall. The Effect of Parameter Estimation on Upper-sided Bernoulli Cumulative Sum Charts. *Quality and Reliability Engineering International*, 29(5):639–651, 2013.

Z. Li and Z. Wang. Adaptive CUSUM of the Q Chart. *International Journal of Production Research*, 48(5): 1287–1301, 2010.

Z. Li, J. Zhang, and Z. Wang. Self-Starting Control Chart for Simultaneously Monitoring Process Mean and Variance. *International Journal of Production Research*, 48(15):4537–4553, 2010.

L. Liu, X. Zi, J. Zhang, and Z. Wang. A Sequential Rank-based Nonparametric Adaptive EWMA Control Chart. *Communications in Statistics-Simulation and Computation*, 42(4):841–859, 2013.

L. Liu, B. Chen, J. Zhang, and X. Zi. Adaptive Phase II Nonparametric EWMA Control Chart with Variable Sampling Interval. *Quality and Reliability Engineering International*, 31(1):15–26, 2015.

M.T. Longnecker and T.P. Ryan. A Deficiency for Residuals Charts for Correlated Data. Technical report, Department of Statistics, Texas A & M, 1991.

C.W. Lu and M.R. Reynolds Jr. Ewma Control Charts for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology*, 31(2):166, 1999.

C.W. Lu and M.R. Reynolds Jr. CUSUM Charts for Monitoring an Autocorrelated Process. *Journal of Quality Technology*, 33(3):316, 2001.

J.M. Lucas and M.S. Saccucci. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32(1):1–12, 1990.

T. Lwin. Parameter Estimation in First-Order Autoregressive Model for Statistical Process Monitoring in the Presence of Data Autocorrelation. *Journal of Statistical Planning and Inference*, 141(8):2556–2575, 2011.

E.M. Maboudou-Tchao and D.M. Hawkins. Self-Starting Multivariate Control Charts for Location and Scale. *Journal of Quality Technology*, 43(2):113–126, 2011.

P. Maravelakis and P. Castagliola. An EWMA Chart for Monitoring the Process Standard Deviation when Parameters are Estimated. *Computational Statistics & Data Analysis*, 53(7):2653–2664, 2009.

P.E. Maravelakis, J. Panaretos, and S. Psarakis. Effect of Estimation of the Process Parameters on the Control Limits of the Univariate Control Charts for Process Dispersion. *Communications in Statistics – Simulation and Computation*, 31(3):443–461, 2002.

F.H.C Marriot and J.A. Pope. Bias in the Estimation of Autocorrelation. *Biometrika*, 41(3-4):390–402, 1954.

A.M. Mathai and S.B. Provost. *Quadratic Forms in Random Variables: Theory and Applications*. Dekker, 1992.

A.K. McCracken and S. Chakraborti. Control Charts for Joint Monitoring of Mean and Variance: An Overview. *Quality Technology & Quantitative Management*, 10(1):17–36, 2013.

D.C. Montgomery. *Introduction to Statistical Quality Control*. John Wiley & Sons, 2007.

R. Osei-Aning, S.A. Abbasi, and M. Riaz. Mixed EWMA-CUSUM and mixed CUSUM-EWMA Modified Control Charts for Monitoring First Order Autoregressive Processes. *Quality Technology & Quantitative Management*, 14(4):429–453, 2017a.

R. Osei-Aning, S.A. Abbasi, and M. Riaz. Optimization design of the cusum and ewma charts for autocorrelated processes. *Quality and Reliability Engineering International*, 33(8):1827–1841, 2017b.

E.S. Page. Continuous Inspection Schemes. *Biometrika*, pages 100–115, 1954.

D.R. Prajapati and S. Singh. Control Charts for Monitoring the Autocorrelated Process Parameters: A Literature Review. *International Journal of Productivity and Quality Management*, 10(2):207–249, 2012.

S.B. Provost and D. Sangel. Inference about the First Order Autoregressive Coefficient. *Communications in Statistics - Theory and Methods*, 34(5):1183–1201, 2005.

S. Psarakis and G.E.A. Papaleonida. Spc Procedures for Monitoring Autocorrelated Processes. *Quality Technology and Quantitative Management*, 4(4):501–540, 2007.

S. Psarakis, A.K. Vyniou, and P. Castagliola. Some Recent Developments on the Effects of Parameter Estimation on Control Charts. *Quality and Reliability Engineering International*, 30(8):1113–1129, 2014.

M.H. Quenouille. Notes on Bias in Estimation. *Biometrika*, 43(3-4):353–360, 1949.

C.P. Quesenberry. SPC Q Charts for Start-Up Processes and Short or Long Runs. *Journal of Quality Technology*, 23(3):213–224, 1991a.

C.P. Quesenberry. SPC Q Charts for a Binomial Parameter p: Short or Long Runs. *Journal of Quality Technology*, 23(3):239–246, 1991b.

C.P. Quesenberry. SPC Q Charts for a Poisson Parameter λ: Short or Long Runs. *Quality control and applied statistics*, 37(3):117–120, 1992.

M.R. Reynolds, R.W. Amin, J.C. Arnold, and J.A. Nachlas. $\bar{X}$ Charts with Variable Sampling Intervals. *Technometrics*, 30(2):181–192, 1988.

S.W. Roberts. Control Chart Tests based on Geometric Moving Averages. *Technometrics*, 1(3):239–250, 1959.

G.C. Runger and T.R. Willemain. Model-Based and Model-Free Control of Autocorrelated Processes. *Journal of Quality Technology*, 27(4):283–292, 1995.

G.C. Runger and T.R. Willemain. Batch-means Control Charts for Autocorrelated Data. *IIE transactions*, 28 (6):483–487, 1996.

N.A. Saleh, M.A. Mahmoud, L.A. Jones-Farmer, I. Zwetsloot, and W.H. Woodall. Another Look at the EWMA Control Chart with Estimated Parameters. *Journal of Quality Technology*, 47(4):363–382, 2015a.

N.A. Saleh, M.A. Mahmoud, M.J. Keefe, and W.H. Woodall. The Difficulty in Designing Shewhart $\bar{X}$ and X Control Charts with Estimated Parameters. *Journal of Quality Technology*, 47(2):127–138, 2015b.

N.A. Saleh, I.M. Zwetsloot, M.A. Mahmoud, and W.H. Woodall. Cusum Charts with Controlled Conditional Performance under Estimated Parameters. *Quality Engineering*, 28(4):402–415, 2016.

A. Schöne and W. Schmid. On the Joint Distribution of a Quadratic and a Linear Form in Normal Variables. *Journal of multivariate analysis*, 72(2):163–182, 2000.

L.R. Shenton and W. Johnson. Moments of a Serial Correlation Coefficient. *Journal of the Royal Statistics Society - Series B (Methodological)*, 27(2):308–320, 1965.

W.A. Shewhart. *Economic Control of Quality of Manufactured Product*, volume 509. ASQ Quality Press, 1931.

A. Snoussi, M.E. Ghourabi, and M. Limam. On SPC for Short Run Autocorrelated Data. *Communications in Statistics - Simulation and Computation*, 34(1):219–234, 2005.

A.P. Stamboulis. First Order Autoregressive Model Applied to Quality Control. *Unpub. Memorandum. New York Univ., New York*, 1971.

J.H. Sullivan and L.A. Jones-Farmer. A Self-Starting Control Chart for Multivariate Individual Observations. *Technometrics*, 44(1):24–33, 2002.

W.L. Teoh, M.B.C. Khoo, P. Castagliola, and S. Chakraborti. Optimal Design of the Double Sampling $\bar{X}$ Chart with Estimated Parameters Based on Median Run Length. *Computers & Industrial Engineering*, 67:104–115, 2014.

K. Triantafyllopoulos and S. Bersimis. Phase II Control Charts for Autocorrelated Processes. *Quality Technology & Quantitative Management*, 13(1):88–108, 2016.

A.K. Tsui and M.M. Ali. Exact Distributions, Density Functions and Moments of the Least Squares Estimator in a First Order Autoregressive Model. *Computational Statistics and Data Analysis*, 17(4):433–454, 1994.

A.V. Vasilopoulos and A.P. Stamboulis. Modification of Control Chart Limits in the Presence of Data Correlation. *Journal of Quality Technology*, 10(1):20–30, 1978.

S. Wang. Tail Probability Approximations in the First Order Noncircular Autoregressive. *Biometrika*, 79(2): 431–434, 1992.

C.H. Weiß, D. Steuer, C. Jentsch, and M.C. Testik. Guaranteed Conditional ARL Performance in the Presence of Autocorrelation. *Computational Statistics & Data Analysis*, 128:367–379, 2018.

D. J. Wheeler. Rational subgrouping: The conceptual foundation of process behavior charts. *Quality Digest Daily*, 2015. (retrieved from `https://www.spcpress.com/pdf/DJW282.pdf`).

Y. Wu. Inference for Post-Change Parameters After Sequential CUSUM Test Under AR(1) Model. *Journal of Statistical Planning and Inference*, 168:52–67, 2016.

W.C. Yeong, M.B.C. Khoo, Y. Ou, and P. Castagliola. Economic-Statistical Design of the Synthetic $\bar{X}$ Chart with Estimated Process Parameters. *Quality and Reliability Engineering International*, 31(5):863–876, 2015.

P.F. Zantek. Design of Cumulative Sum Schemes for Start-Up Processes and Short Runs. *Journal of Quality Technology*, 38(4):365–375, 2006.

C.W. Zhang, M. Xie, and T. Jin. An Improved Self-Starting Cumulative Count of Conforming Chart for Monitoring High-Quality Processes under Group Inspection. *International Journal of Production Research*, 50 (23):7026–7043, 2012.

M. Zhang, Y. Peng, A. Schuh, F.M. Megahed, and W.H. Woodall. Geometric Charts with Estimated Control Limits. *Quality and Reliability Engineering International*, 29(2):209–223, 2013.

M. Zhang, F. Megahed, and W.H. Woodall. Exponential CUSUM Charts with Estimated Control Limits. *Quality and Reliability Engineering International*, 30(2):275–286, 2014.

N.F. Zhang. A Statistical Control Chart for Stationary Process Data. *Technometrics*, 40(1):24–38, 1998.

N.F. Zhang and A.L. Pintar. Monitoring Process Variability for Stationary Process Data. *Quality and Reliability Engineering International*, 31(8):1383–1396, 2015.

M.J. Zhao and A.R. Driscoll. The c-Chart with Bootstrap Adjusted Control Limits to Improve Conditional Performance. *Quality and Reliability Engineering International*, 32(8):2871–2881, 2016.

C. Zhou, C. Zou, Y. Zhang, and Z. Wang. Nonparametric Control Chart Based on Change-Point Model. *Statistical Papers*, 50(1):13–28, 2009.

M. Zhou and T.N. Goh. Effects of Model Accuracy on Residual Control Charts. *Quality and Reliability Engineering International*, 32(5):1785–1794, 2016.

C. Zou, C. Zhou, Z. Wang, and F. Tsung. A Self-Starting Control Chart for Linear Profiles. *Journal of Quality Technology*, 39(4):364, 2007.

I.M. Zwetsloot and W.H. Woodall. A Head-to-Head Comparative Study of the Conditional Performance of Control Charts based on Estimated Parameters. *Quality Engineering*, 29(2):244–253, 2017.

I.M. Zwetsloot, M. Schoonhoven, and R.J.M.M. Does. A Robust Phase i Exponentially Weighted Moving Average Chart for Dispersion. *Quality and Reliability Engineering International*, 31(6):989–999, 2015.

Published Papers

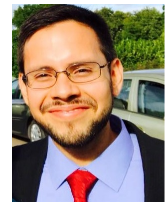
1. J.A. Garza-Venegas, V.G. Tercero-Gómez, A. Cordero-Franco, M. Temblador-Pérez, and M. Beruvides. An Evaluation of Change-point Estimators for a Sequence of Normal Observations with Unknown Parameters. *Communications in Statistics-Simulation and Computation*, 46(6):4297–4317, 2017. doi: 10.1080/03610918.2015.1115524

2. J.A. Garza-Venegas, V.G. Tercero-Gómez, L. Lee Ho, P. Castagliola, and G. Celano. Effect of Autocorrelation Estimators on the Performance of the $\bar{X}$ Control Chart. *Journal of Statistical Computation and Simulation*, 88(13):2612–2630, 2018. doi: 10.1080/00949655.2018.1479752

Curriculum Vitae

Jorge Garza

Address: *Marsella 149 Col. Santaluz, General Escobedo, Nuevo León, México*
Mobile: *(044) 811 170 5938*
Phone: *01 81 1493 1318*
E-mail: *jag.venegas@gmail.com*



Education

- **2006–2010: Bachelor of Mathematics**, *Facultad de Ciencias Físico - Matemáticas, Universidad Autónoma de Nuevo León*, San Nicolás de los Garza, Nuevo León, México. GPA – 9.2.
- **2011–2013: Master of Science in Mathematics**, *Facultad de Ciencias Físico - Matemáticas, Universidad Autónoma de Nuevo León*, San Nicolás de los Garza, Nuevo León, México. GPA – 9.5.
  Thesis title: “Analysis of multiple change-points in normally distributed series”.
- **2015–2018: Ph.D in Engineering Sciences**, *School of Engineering and Sciences, Tecnologico de Monterrey*, Monterrey, Nuevo León, México. GPA – 9.7.
  Thesis title: “Control charts for autocorrelated processes under parameter estimation”.

Work Experience

- **Jan - Jun 2018**, FACULTAD DE AGRONOMÍA, UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN.
  Courses taught (undergraduate courses): Differential Calculus, Differential Equations, Probability and Statistics.
- **Jan - Jun 2016**, FACULTAD DE INGENIERÍA CIVIL, UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN.
  Courses taught (undergraduate courses): Differential Equations.
- **Aug 2012 - Dec 2014**, FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS, UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN.
  Courses taught (undergraduate courses): Differential Calculus, Algebra, Matrix Algebra, Linear Algebra, Calculus of Several Variables, Differential Equations, Discrete Mathematics, Statistics, Complex Analysis, Group Theory, Real Analysis.
- **Jan - Aug 2012**, UNIVERSIDAD METROPOLITANA DE MONTERREY.
  College teacher.
- **Feb 2010 - Dec 2012**, CASA PATERNA “LA GRAN FAMILIA”
  Volunteer. Advisor of residents of the children’s home.

This document was typed in using LaTeX by Jorge Arturo Garza Venegas.