Università
degli Studi
di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXV

GENERAL PROFILE MONITORING THROUGH NONPARAMETRIC TECHNIQUES

Direttore della Scuola: Prof.ssa MONICA CHIOGNA

Supervisore: Prof. GUIDO MASAROTTO

Co-supervisore: Prof.ssa GIOVANNA CAPIZZI

Dottorando: ENRICO TONINI

31/01/2013

Ad Alessandra

Acknowledgements

First of all, for its particular controversial nature it has had for me, I would like to dedicate this work to my father Oreste, who has gone some years ago. I would like him a lot to see all results and experiences of my life I have been getting in the last years.

I would like to thank my family, in particular my mother Maria, Franz and my brother Alessandro, for their continuous, precious and moral support during these seven and half years. I am saying seven and half and not three years, because I like thinking these Ph.D. period as the conclusion of a wonderful and wealthy career in the Faculty and in the Department of Statistics in the University of Padova consisting also in Bachelor and Master Degree in Statistics.

A second great thank you is due to my supervisor, Professor Guido Masarotto. Without his precious help and encouragements I would have never been able to produce this work. I like to say that I regard him very highly as a professor, as a teacher, but most of all as a person.

I thank my cosupervisor, Professor Giovanna Capizzi, for her support in this work too.

I would like to thank my friends and colleagues, who have become very special friends for me. I think I will never be able to forget the collaboration and the friendship which we begun almost immediately to establish in a very natural and special way. In particular I will always keep a special place in my heart for Manuela, Ivan, Clovis and Federico, which have been even more than colleagues and special friends for me and with whom I will always try to keep the nice friendship struck up in these three years.

I would like to thank also all the other friends and relatives of mine which have been very close to me in these three years of Ph.D., in particular Samuel, Matteo, Camilla, Lele, Enrico, Mimmo, Beppe, Tommy, Guen, Simone, Elisa, Massi, Naty, il Líder, Andrea, Ele, Pierre, Rick, Katia, Gugli, Krizia, Razzo, Ponz, Bic, Luisa, Detta, Claudio, Chiara, Stefano, Giulio, Alessandra, Martina, Alessandro, Luca, Elisa, Rosso.

Abstract

This Ph.D. thesis is devoted to Statistical Process Control (SPC) methods for monitoring over time the stability of a relation between two variables (profile). Very often in literature the functional form of the relation is assumed to be known, whereas in this work we concentrated on generic and unknown relations which have to be estimated with the usual nonparametric regression techniques. The original contributes are two, presented in chapters 2 and 3 respectively. In Chapter 1 we make a brief overview on the topic in order to make you become familiar with these specific problems of Statistical Process Control (SPC) applications and we introduce you to the original parts of this work. In Chapter 2 we envelope and compare five new control charts for monitoring *on-line* unknown general, and not only linear, relations among variables over time under the assumption of the normality of the errors; these charts combine in an original way the following techniques: self-starting methods, useful to drop the distinction between Phase I and Phase II of the analysis; very known multivariate charting schemes as MEWMA and CUSCORE; nonparametric testing techniques as wavelet methods and kernel linear smoothing. In Chapter 3, instead, we construct a test statistic useful to check with a completely nonparametric procedure the stability of a process retrospectively, thus *off-line*. Both second and third chapters are structured in the following way: brief literature review; framework and model considered in our study; simulation study; a section with some useful complements on the topics and relative research carried out; conclusion and suggestions for future research.

Sommario

Questa tesi è dedicata ai metodi per il Controllo Statistico della Qualità (CSQ) per il monitoraggio della stabilità nel tempo della relazione tra due variabili (profilo). Spesso in letteratura si assume nota la forma funzionale della relazione, viceversa in questo lavoro ci si è concentrati su relazioni generiche ed ignore e quindi da stimare con le usuali tecniche di regressione non parametrica. I contributi originali sono due, presentati nei capitoli 2 e 3 respectively. Nel Capitolo 1 presentiamo una breve panoramica dell'argomento in modo da far prendere familiarità al lettore con questi problemi specifici delle applicazioni del Controllo Statistico della Qualità (CSQ) e introdurlo alle parti originali di questo lavoro. Nel Capitolo 2 sviluppiamo e confrontiamo cinque nuove carte di controllo per il monitoraggio *on-line* di relazioni ignote generiche, e non solo lineari, tra variabili sotto l'assunzione di normalità degli errori; queste carte mettono insieme in modo originale le seguenti tecniche: metodi self-starting, utili per eliminare la distinzione tra Fase I e Fase II dell'analisi; alcune carte di controllo multivariate ben note come MEWMA e CUSCORE; tecniche non parametriche per la verifica di ipotesi come metodi wavelet o il lisciamento lineare con il metodo del kernel. Nel Capitolo 3, invece, costruiamo una statistica test utile per verificare con una procedura completamente non parametrica la stabilità di un processo in maniera retrospettiva, quindi *off-line*. Sia il secondo che il terzo capitolo sono strutturati nel modo seguente: breve revisione della letteratura; contesto e modello considerati in questo studio; simulazioni; una sezione con alcuni complementi utili sugli argomenti e relativa ricerca effettuata; conclusione e suggerimenti per la ricerca futura.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

In most Statistical Process Control (SPC) applications, it is assumed that the quality of a process or product can be adequately represented by the distribution of a univariate quality characteristic or by the general multivariate distribution of a vector consisting of several correlated quality characteristics. In many practical situations, however, the quality of a process or product is better characterized and summarized by a relation between a response variable and one or more explanatory variables. Profile monitoring is used to understand and to check the stability of this relation over time. At each sampling stage, one observes a collection of data points that can be represented by a curve (or profile). In some calibration applications, the profile can be adequately represented by a simple straight-line model, while in other applications more complicated models are needed, involving nonlinear profiles and the use of nonparametric techniques.

Lots of interesting works about profile monitoring are summarized in two useful review papers (Woodall et al., 2004; Woodall, 2007) and in a recent book (Noorossana et al., 2011). Most of work has been focused in linear profile monitoring, whereas less work has been done in nonlinear and even less in general/nonparametric profiles.

In this profile monitoring framework, the so-called Self-Starting (SS) control schemes (see for example Hawkins & Maboudou-Tchao (2007) and Capizzi & Masarotto (2010b)) are very useful; they consist in transforming the original data in a proper way such that we get rid of the estimation of the parameters of the In-Control (IC) process: this is a good attempt to update traditional control schemes, which are usually designed with the assumption that the IC parameters are exactly known. Self-starting control schemes, instead, drop the traditional distinction between a retrospective analysis phase (Phase I), where one has to be very accurate to check statistical control and establish control limits, and a prospective monitoring phase (Phase II) for testing process stability as new samples are collected, and this allows us to avoid adding a random element (the estimation of the IC parameters in Phase I) to the Run Length (RL) distribution, the random variable describing when a control chart signals an out-of-control situation, and to use all the data immediately to update the parameter estimates and simultaneously check for Out-of-Control (OC) conditions.

In general for Phase I lots of IC observations are needed to estimate accurately the parameters of the IC distribution of the process, a fortiori in presence of a profile, where even more parameters are unknown, in particular when we do not know its shape and we use a nonparametric approach. Therefore, self-starting control schemes are favored in particular: (1) when early OC production is costly; (2) when there is considerable delay between production units; and (3) when samples sufficiently large to approximate control chart performance with the true parameters are unavailable.

Therefore in the first part of this thesis we want to consider only self-starting charting schemes and we propose and compare five new control charts for monitoring general, and not only linear, relations among variables over time; these charts combine self-starting methods, very known multivariate charting schemes as MEWMA and CUSCORE and nonparametric testing techniques. This work refers to general profiles and its original idea is just to try to use together self-starting charting schemes to get rid of the in-control parameters and nonparametric techniques, such as wavelet transforms and kernel linear smoothing, to synthesize the information about the relation among the response and the explanatory variables.

Self-starting control charts are thus very useful, but sometimes it is necessary to keep the distinction between Phase I and Phase II and in SPC it is crucial to check the stability of a process in Phase I. Even here, some methods which do this in case of a known relation (linear or not linear) are already existing. Therefore, along the lines of the first part, we would like to propose a new method completely nonparametric, able to assess the stability of a general and unknown relation among variables. This new method is actually a multivariate version of an already existing method, recently proposed by Capizzi & Masarotto (2012), which tries to make us realize if, in Phase I, a general profile is stable or not. Furthermore this method will provide an interesting statistical tool to make some diagnostics in case of instability of the process, that is to try to understand where the process has started to go out of control.

1.2 Main Contribution of the Thesis

In our first study the reference model used for the profile data in our first study is the following simple multivariate Gaussian change-point model:

$$\boldsymbol{y}_t = \mu(\boldsymbol{x}) + \delta(\boldsymbol{x})\mathbb{I}_{[\tau,+\infty)}(t) + \boldsymbol{\epsilon}_t\,, \quad t = 1, 2, \ldots\,,$$

where $\mu(\boldsymbol{x})$ and $\delta(\boldsymbol{x})$ are smooth functions representing the In-Control (IC) profile and its Out-of-Control (OC) shift occurring from the τ^{th} observation on, and $\boldsymbol{\epsilon}_t \sim \mathrm{N}_n(\mathbf{0}_n, \sigma^2 \boldsymbol{I}_n)$ is the error term of the model. After a proper self-starting standardization of the profile data, such that we know the IC distribution of this transformation, we compared five self-starting charting schemes:

- Self-Starting MEWMA (SSMEWMA);
- Self-Starting MEWMA chart with Wavelet thresholding Fan's test (SS-WFMEWMA);
- Self-Starting CUSCORE chart (SSCUSCORE);
- Self-Starting CUSCORE chart with Wavelet thresholding Fan's test (SSWFCUSCORE);
- Self-Starting MEWMA chart with kernel Nonparametric smoothing (SSNEWMA).

We investigated the monitoring performance of the five proposed profile monitoring schemes and their parameters in a simulation study which considers one IC model and three different groups of OC models (all together 90 OC models), which consider a wide variety of types of deviations (shifted, oscillatory and local deviations). Actually, the first control chart, SSMEWMA, is already existing, since it is simply the Multivariate EWMA (MEWMA) arranged to profile data; thus it is more correct to say that performances of the other four control schemes are compared to performances of SSMEWMA.

The performance of the charts in the different scenarios is evaluated through the Relative Mean Index (RMI), which measures the average relative efficiency in terms of Average Run Length (ARL) for a range of shift sizes; a control chart with a smaller RMI is considered better in its overall performance.

In general the nonparametric charting scheme using kernel linear smoothing, SSNEWMA, seems to be the best one, since it is the only one which behaves almost always better than the other ones, whereas wavelet transforms do not seem very performing. In a second moment we tried also to construct an adaptive version of the SSNEWMA chart which could be more performing, but we saw that it is indeed not substantially better than it.

For the second purpose of this thesis, we considered a model more general than model used before:

$$\boldsymbol{y}_t = \mu(\boldsymbol{x}) + \delta(\boldsymbol{x})\mathbb{I}_{[\tau_1,\tau_2]}(t) + \boldsymbol{\epsilon}_t\,, \quad t = 1,\ldots,T\,,$$

where differently from before, the shift $\delta(\boldsymbol{x})$ occurs from the τ_1^{th} profile to the τ_2^{th} one and $\ell = \tau_2 - (\tau - 1) = \tau_2 - \tau_1 + 1$ is the length of the instability period of the process; in this framework we have to establish a number of observations T and $\boldsymbol{\epsilon}_t$, is a quite general multivariate error term, considering also the possibility of some forms of intra-profiles correlation (for further details see model 3.5 at page 49).

The method we propose considers, for each possible combination of begin and end points where the profile may be not stable, a proper beginning statistic, the vector of the sample difference of the means between the unstable and the stable intervals, which is then smoothed and standardized. The final statistic is the maximization of this statistic with respect to all possible instability points and the degrees of freedom of the considered smoother. To compute a p-value which should give us a strong indication about the stability (H_0) or instability (H_1) of a process, we use a permutation method which essentially exchanges the profiles and computes the previously described statistic for each permutation and then compares the value of the statistic computed on the data with the values of the statistic computed on the permutations of the data.

The method proposed in the second part of this thesis to establish the stability of a process is also able to make some diagnostics on the data, in order to discover where (and not only if) an instability period in the process has occurred.

Chapter 2

Self-Starting Control Charts for Monitoring General Profiles Using Nonparametric Techniques

2.1 A Brief Review on Profile Monitoring

In Statistical Process Control (SPC) literature the investigation of the topic of profile monitoring is quite recent. Differently from simpler and more common SPC problems, in which we want to monitor one or more parameters of a univariate or a multivariate distribution which is assumed to describe completely the process, profile monitoring consists in monitoring a whole relation among variables. This problem is crucial for lots of production processes because very often the quality of a process or product can be better represented through a relation between a response variable and some explanatory variables instead of simply one or more parameters of a univariate or a multivariate distribution. Lots of interesting works about profile monitoring are summarized in two useful review papers about profile monitoring (Woodall et al., 2004; Woodall, 2007) and in a recent book by Noorossana et al. (2011); furthermore Colosimo & Pacella (2008) compared different approaches already present in SPC literature to monitor profiles.

Most of work has been done when the structure of the relation among variables is known, in particular when it is linear (linear profile monitoring) and this substantially consists in finding a charting scheme to monitor the regression coefficients (and possibly the variance of the error term) of a linear regression.

The first works about linear profile monitoring suggest for Phase I (Stover & Brill, 1998) and Phase II (Kang & Albin, 2000) methods to monitor the coefficients of a simple linear regression, based on Hotelling's T^2. Kim et al. (2003) improved the previous methods by adding also an EWMA scheme to monitor the variance of the error term and not only the regression coefficients. There are lots of other papers who investigated accurately the monitoring of the regression coefficients: Shu et al. (2004) proposed two control charts (based on Shewhart's and EWMA schemes) with estimated parameters for monitoring the residuals of the regression; Zou et al. (2006) and Mahmoud et al. (2007) proposed two change-point methods to detect possible changes in the regression coefficients through Likelihood Ratio (LR) or/and the EWMA scheme. Other proposals have been suggested by Zou et al. (2007a), who proposed a method based on a MEWMA scheme, by Akhavan Niaki et al. (2007), whose method is based on the Generalized Linear Test (GLT) and the R chart for monitoring the regression coefficients and the variance of the error term, and by Zou et al. (2007b), who proposed a self-starting control charts for monitoring linear profiles. There are also other proposals to monitor linear profiles, which are even more recent, such as Saghaei et al. (2009), who proposed a method based on CUSUM scheme that improves Phase II, Zhang et al. (2009), who proposed another method based on LR, Zou & Qiu (2009), Zou et al. (2010) and Capizzi & Masarotto (2011), who suggested methods based on variables selection.

But experience teaches that very often relations among variables are not linear, nor cannot be linearized at all. There are, indeed, some papers who refer their analysis to nonlinear profiles: Brill (2001) suggested how to use a Hotelling's T^2-based method on the estimates of the coefficients of a nonlinear regression; Williams et al. (2007) proposed extensions of this method by suggesting different estimates of the covariance matrix of the regression coefficients; Mosesova et al. (2006) and Jensen & Birch (2009) proposed the use of Generalized Linear Mixed Models (GLMM) to deal with nonlinear profiles; moreover, some works deal with nonlinear profiles by using wavelet transforms (Chicken et al., 2009).

There are also some authors working on the branch of profile monitoring who enveloped some nonparametric methods (Zou et al., 2009; Qiu & Zou, 2010), sometimes applied to linear profiles where no assumption on distribution of the error term is made (Zi et al., 2011); in particular we signal Zou et al. (2008), where a nonparametric version of the MEWMA scheme, the so-called NEWMA, is proposed: we will exploit this technique among the methods we will propose in the first part of this thesis.

Of particular interest there are charting schemes which use wavelet transforms (see Nason (2008) to know more about them and Zeisset (2008) to see a review of wavelet methods in SPC) to concentrate the basic information of a profile in few parameters: Jin & Shi (1999) proposed some Shewhart's control schemes for monitoring changes in wavelet coefficients; Chicken et al. (2009) suggested a method for monitoring nonlinear profiles using wavelet transforms through a semiparametric innovative approach. Wavelet methods are often used with thresholding, in order to highlight the main noise in a curve or, more in general, in a functional relation among variables. The first proposals in this sense are those made in Fan (1996) and Fan & Lin (1998), whose results are then drawn in Jeong & Lu (2006), where more proposals of "Fan's test" are presented. These so-called Fan's tests should be capable to reduce the in-control variability of the process, by deleting useless noise, without reducing its out-of-control variability. This is why we think they could better signal weak signals than traditional tests, like Hotelling's T^2.

Furthermore, as we have already said in the Overview of this thesis, in the first part of this work we are also particularly focused on self-starting control charts. A first example of a SS scheme in SPC literature comes from Hawkins (1987), where a self-starting CUSUM chart for monitoring the mean and the variance of univariate normal observations is proposed; the topic was drawn in other articles, but in particular we highlight Sullivan & Jones (2002), who suggested a MEWMA approach in case of multivariate normal observations for monitoring the mean vector, and Hawkins & Maboudou-Tchao (2007), who improved this approach with some further tricks; Capizzi & Masarotto (2010b), instead, proposed a CUSCORE chart to get even better performances in monitoring the mean vector of multivariate normal distributions. To this day there are, instead, few examples in literature of self-starting charts used to monitor profiles and not simply parameters: Zou et al. (2007b) presented a self-starting chart for monitoring simple linear profiles, useful for detecting shifts in the intercept, the slope and the standard deviation of the error term; in Qiu & Zou (2010) nonparametric profile monitoring with arbitrary design points was investigated and the authors gave also a self-starting version of the solution.

You can find a good review on profile monitoring with lots of the previous methods in a book (Noorossana et al., 2011).

In this chapter we will try to combine efficiently the techniques already existing in literature in order to obtain new flexible charting schemes which could be possibly useful in as many situations as possible; therefore we want to try to combine the efficiency of the self-starting charts, in order to use all the data and not to distinguish between the two phases of the analysis, and the ability of nonparametric techniques, such as wavelets transforms and kernel linear smoothing, to synthesize a relation among variables and to let us free not to hypothesize the form of the profile, that is the type of relation between the response and the explanatory variables.

This chapter is arranged in the following way: in Section 2.2 we will expose the reference model and our proposed charts to apply to it, clearly distinguishing the common self-starting part and the five ways to accumulate the observations and obtain the test statistic; in Section 2.3 we will present our simulation study to test our proposed control schemes and we will make a summary of what the results suggest; in Section 2.4 we will give some useful complements on the topic and will present other research tools carried out; finally, in Section 2.5 we will try to give some hints for future research about this topic to improve results obtained in this first part of the thesis.

2.2 Framework and Model

In order to build a control chart to monitor profiles (or in general to monitor the parameters of any process over time), we need to make three proper choices and combine them:

1. how to deal with unknown parameters in the IC process;
2. how to accumulate the profiles over time;
3. how to accumulate the profiles with respect to the n observations and thus to construct a test statistic which will try to highlight OC signals.

To deal with the unknown parameters of the IC process, we will use in any case self-starting control charts. These charts have the great advantage to remove the dependence of the data on unknown parameters; self-starting charts, indeed, allow to transform the data such that the transformation performed has a known IC distribution. They are efficient in the sense they do not need the plug-in of the estimates of the true parameters and to force us to distinguish between a I phase, where we collect lots of data in order to estimate the unknown parameters, and a II phase, where we apply the chosen chart with estimated parameters to other data.

To accumulate the observations over time we use either a Multivariate Exponentially Weighted Moving Average charting scheme (MEWMA) or a CUSCORE charting scheme (see Montgomery (2005) for an excellent documentation on the most important charting schemes).

The test statistics considered are either the typical statistics of CUSCORE or MEWMA charts, or Hotelling's T^2; in some cases, these statistics are properly modified by applying them the kernel linear smoothing or a wavelet transform (Nason, 2008) with a thresholding Fan's test (Fan & Lin, 1998).

By properly combining these choices, we consider five types of control charts, which we will try to use on simulated profile data with respect to different OC scenarios with small values of the global OC parameter for comparing their performances in detecting small deviations from a IC situation (see Section 2.3):

1. self-starting methods + MEWMA chart + typical MEWMA chart test statistic: Self-Starting Multivariate Exponentially Weighted Moving Average chart (SSMEWMA);

2. self-starting methods + MEWMA chart + Wavelet thresholding Fan's test: Self-Starting Multivariate Exponentially Weighted Moving Average with Wavelet thresholding Fan's test chart (SSWFMEWMA);

3. self-starting methods + CUSCORE chart + typical CUSCORE chart test statistic: Self-Starting CUSCORE chart (SSCUSCORE);

4. self-starting methods + CUSCORE chart + Wavelet thresholding Fan's test: Self-Starting CUSCORE with Wavelet thresholding Fan's test chart (SSWFCUSCORE);

5. self-starting methods + MEWMA chart + Kernel: Self-Starting Nonparametric Multivariate Exponentially Weighted Moving Average (SSNEWMA).

Now let us introduce the common part to which we apply the previous five charting schemes, which are the reference model of the profile data on which we will work and the self-starting standardization of these data. Consistently with recent literature about self-starting charts (Sullivan & Jones, 2002; Capizzi & Masarotto, 2010b), the general model we will refer to present and study the performances of our charting schemes in nonparametric profile monitoring is the following simple multivariate Gaussian change-point model:

$$\boldsymbol{y}_t = \mu(\boldsymbol{x}) + \delta(\boldsymbol{x})\mathbb{I}_{[\tau,+\infty)}(t) + \boldsymbol{\epsilon}_t\,, \quad t = 1, 2, \dots\,, \tag{2.1}$$

where:

- $\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \dots, y_{n,t})'$ is an n-dimensional vector representing the response variable of the considered profile; we consider n to be at least "a few dozen", so that the knowledge of $\boldsymbol{y}_t$ and $\boldsymbol{x}$ can represent in a quite good way the relation between the two variables at time t;

- $\mu(\boldsymbol{x}) = (\mu(x_1), \mu(x_2), \dots, \mu(x_n))'$, independent of time, is a (sufficiently) smooth function which represents the relation between the response variable $\boldsymbol{y}$ and the deterministic explanatory variable $\boldsymbol{x}$; we therefore reduce our study to the case of only one explanatory variable;

- $\delta(\boldsymbol{x}) = (\delta(x_1), \delta(x_2), \dots, \delta(x_n))'$ is a (sufficiently) smooth function which represents the deviation from the IC model occurred after $(\tau - 1)$ time units;

- $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{n,t})' \sim \mathrm{N}_n(\mathbf{0}_n, \sigma^2\boldsymbol{I}_n)$ is the error term of the model; we therefore reduce our study to the case of independent, Gaussian and homoschedastic errors;

- $\mathbb{I}_A(x)$ is the indicator function of x with respect to the set A, therefore

$$\mathbb{I}_{[\tau,+\infty)}(t) = \begin{cases} 1 & \text{if } t \geq \tau \\ 0 & \text{if } t < \tau\,. \end{cases}$$

Before applying our five control charts, in order to get rid of the unknown parameters, which are, in this case, the mean function $\mu(\cdot)$ and the variance of the error σ^2, we apply a self-starting standardization to the profile data similar to those presented in Sullivan & Jones (2002) and Hawkins & Maboudou-Tchao (2007) and adapted to the case of our model.

The first step of this standardization is

$$\boldsymbol{b}_t = a_t(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{t-1}), \quad t = 2, 3, \dots,$$

where $a_t = \left(\frac{t-1}{t}\right)^{\frac{1}{2}}$ is a standardizing constant and $\overline{\boldsymbol{y}}_{t-1} = \frac{1}{t-1}\sum_{j=1}^{t-1}\boldsymbol{y}_j$ is the observation mean vector of the first $(t-1)$ profiles. This transformation allows us to get rid of the mean function $\mu(\cdot)$ and we have that

$$\boldsymbol{b}_t \sim \mathrm{N}_n(\boldsymbol{0}_n, \sigma^2\boldsymbol{I}_n)$$

when the process is IC at time t, which means that $b_{i,t}$ are independent $\mathrm{N}(0, \sigma^2)$ $\forall\ i = 1, \dots, n$ and $t = 2, 3, \dots$ Then we have to estimate σ^2 to get rid of it and we propose the following estimate:

$$s_t^2 = \frac{1}{2n(t-1)}\sum_{i=1}^{n}\sum_{j=2}^{t}(y_{i,j} - y_{i,j-1})^2, \quad t = 2, 3, \dots. \tag{2.2}$$

We prefer this estimate to other estimates of σ^2 proposed in the literature (Sullivan & Jones, 2002) because it introduces possible bias only between times $\tau - 1$ and τ, that is only at time of the occurred shift in the curve. At this point, to get rid of the dependence on σ^2, we divide $\boldsymbol{b}_t$ by s_{t-1}, getting

$$\boldsymbol{d}_t = \frac{\boldsymbol{b}_t}{s_{t-1}} = \frac{a_t(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{t-1})}{s_{t-1}} = \frac{\left(\frac{t-1}{t}\right)^{\frac{1}{2}}(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{t-1})}{s_{t-1}}, \quad t = 3, 4, \dots.$$

Since $\boldsymbol{b}_t$ and s_{t-1}^2 are approximately independent we have that

$$\boldsymbol{d}_t \mathrel{\dot\sim} \mathrm{T}_{n,n(t-2)}$$

when the process is IC at time t, where $\mathrm{T}_{n,\nu}$ is the multivariate n-dimensional Student's t distribution with ν degrees of freedom. This means that $d_{i,t}$ are approximately independent $\mathrm{t}_{n(t-2)}$ $\forall\ i = 1, \dots, n$ and $t = 3, 4, \dots$ By using s_{t-1}^2 and not s_t^2, we reduce the possible, anyway slight, dependence between the numerator and the denominator and in this way the approximation to the Student's t distribution is even better (from some tries we empirically saw that scatter plots of the two terms signal no evidence of dependence and the correlation between the two terms, already very low, becomes even lower by delaying s_t^2). Finally, using the probability integral transform approach (Sullivan & Jones, 2002; Hawkins & Maboudou-Tchao, 2007) we get

$$\boldsymbol{q}_t = \Phi^{-1}\left\{F_{n(t-2)}\left[\boldsymbol{d}_t\right]\right\} = \Phi^{-1}\left\{F_{n(t-2)}\left[\frac{a_t(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{t-1})}{s_{t-1}}\right]\right\}, \quad t = 3, 4, \dots, \tag{2.3}$$

where $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution and $F_\nu(\cdot)$ is the distribution function of a Student's t distribution with ν degrees of freedom. An approximated IC distribution for $\boldsymbol{q}_t$ at time t is

$$\boldsymbol{q}_t \dot{\sim} \mathrm{N}_n(\boldsymbol{0}_n, \boldsymbol{I}_n),$$

which means that $q_{i,t}$ are approximately independent $\mathrm{N}(0,1)$ $\forall\ i = 1, \dots, n$ and $t = 3, 4, \dots$ Notice that with n equal to "a few dozen" and already for a small value of t, we have that $F_{n(t-2)}(\cdot) \doteq \Phi(\cdot)$, therefore

$$\boldsymbol{q}_t \doteq \boldsymbol{d}_t = \frac{a_t(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{t-1})}{s_{t-1}}.$$

The previous steps are common to all five charts; now, depending on the combination of the choices we make about accumulating profiles with respect to time and to the n observations for every time t, we get the following charts.

SSMEWMA first applies a MEWMA scheme (Montgomery, 2005) to $\boldsymbol{q}_t$:

$$\boldsymbol{z}_t = (1-\lambda)\boldsymbol{z}_{t-1} + \lambda \boldsymbol{q}_t, \tag{2.4}$$

with $\boldsymbol{z}_0 = \boldsymbol{0}_n$ and $\lambda \in (0,1)$; $\boldsymbol{z}_t$ has an approximated $\mathrm{N}_n\big(\boldsymbol{0}_n, \frac{\lambda}{2-\lambda}\boldsymbol{I}_n\big)$ distribution when the process is IC at time t. Finally this chart computes the usual MEWMA-statistic on $\boldsymbol{z}_t$:

$$\mathrm{SSMEWMA}_t = \frac{2-\lambda}{\lambda}\boldsymbol{z}_t'\boldsymbol{z}_t. \tag{2.5}$$

If the process is IC at time t, $\mathrm{SSMEWMA}_t \dot{\sim} \chi^2_n$ (but they are not independent); the chart signals if $\mathrm{SSMEWMA}_t > h_1$, where h_1 is chosen to achieve a specified IC Average Run Length (ARL_0).

SSWFMEWMA first applies the previous MEWMA scheme (2.4) to $\boldsymbol{q}_t$ and after that it applies a wavelet transform to $\boldsymbol{z}_t$:

$$\boldsymbol{w}_t = \boldsymbol{W}\boldsymbol{z}_t\,,$$

where $\boldsymbol{W}$ is a matrix which defines a particular wavelet transform (Nason, 2008). Since $\boldsymbol{W}$ is an orthonormal matrix, $\boldsymbol{w}_t$ has still an approximated $\mathrm{N}_n\big(\boldsymbol{0}_n, \frac{\lambda}{2-\lambda}\boldsymbol{I}_n\big)$ when the process is IC at time t. Finally this chart computes the usual MEWMA-statistic on $\boldsymbol{w}_t$ and then adjusts it with the typical Fan's thresholding (for further information about Fan's test see (Fan & Lin, 1998)):

$$\mathrm{SSWFMEWMA}_t = \frac{2-\lambda}{\lambda}\sum_{i=1}^{n} w_{i,t}^2 \mathbb{I}_{[k_1,+\infty)}(|w_{i,t}|)\,, \tag{2.6}$$

where $k_1 = \sqrt{\frac{\lambda}{2-\lambda}}k$ is a standardized thresholding constant: to make a good thresholding it is useful to consider high quantiles of the $\mathrm{N}\big(0, \frac{\lambda}{2-\lambda}\big)$ distribution as values for k_1, which means to consider high quantiles of the standard normal distribution as values for k. The wavelet transform should accumulate in the first wavelet coefficients and the threshold keeps only the main wavelet coefficients, that is the main component of the relation between $\boldsymbol{y}$ and $\boldsymbol{x}$. The chart signals if $\mathrm{SSWFMEWMA}_t > h_2$, where h_2 is chosen to achieve a specified ARL_0. Note that if $k = 0$, SSWFMEWMA reduces to SSMEWMA since $\sum_{i=1}^n w_{i,t}^2 = \boldsymbol{w}_t'\boldsymbol{w}_t = \boldsymbol{z}_t'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{z}_t = \boldsymbol{z}_t'\boldsymbol{z}_t$.

SSCUSCORE first applies a recent CUSCORE scheme (Montgomery, 2005) presented by Capizzi & Masarotto (2010b) to $\boldsymbol{q}_t$: starting from $t = 3$, $c_{2,j}^L = c_{2,j}^U = 0$ and $\tau_{2,j}^L = \tau_{2,j}^U = 3$, $j = 1, \ldots, n$, it computes the CUSCORE statistics

$$\begin{aligned} c_{t,j}^L &= \min\left\{0, c_{t-1,j}^L + f_t(\tau_{t-1,j}^L)\left[q_{t,j} + \frac{1}{2}f_t(\tau_{t-1,j}^L)\right]\right\} \\ c_{t,j}^U &= \max\left\{0, c_{t-1,j}^U + f_t(\tau_{t-1,j}^U)\left[q_{t,j} - \frac{1}{2}f_t(\tau_{t-1,j}^L)\right]\right\}\,, \end{aligned}$$

where

$$\tau_{t,j}^L = \begin{cases} t+1 & \text{if } c_{t,j}^L = 0 \\ \tau_{t-1,j}^L & \text{if } c_{t,j}^L < 0\,, \end{cases}$$

$\tau_{t,j}^U$ is analogously defined from $c_{t,j}^U$ and

$$f_t(\tau) = m \max\left\{c, \frac{\tau - 1}{\sqrt{t(t-1)}}\right\} ,$$

where $m > 0$ and $0 < c < 1$ are proper design constants, whose recommended values are 0.25 and 0.5, respectively. The final statistic to monitor is

$$\text{SSCUSCORE}_t = \sum_{i=1}^{n} \max\left(-c_{i,t}^L, c_{i,t}^U\right) . \tag{2.7}$$

The chart signals if $\text{SSCUSCORE}_t > h_3$, where h_3 is chosen to achieve a specified ARL_0.

SSWFCUSCORE applies the same previous CUSCORE scheme to $\boldsymbol{w}_t = \boldsymbol{W}\boldsymbol{q}_t$, a wavelet transform of $\boldsymbol{q}_t$, and the final statistic to monitor is

$$\text{SSWFCUSCORE}_t = \sum_{i=1}^{n} \max\left(-c_{i,t}^L - k, c_{i,t}^U - k, 0\right) , \tag{2.8}$$

where k is a thresholding constant and c_{it}^L and c_{it}^U are computed as in SSCUSCORE scheme, with $\boldsymbol{w}_t$ in place of $\boldsymbol{q}_t$. The chart signals if $\text{SSWFCUSCORE}_t > h_4$, where h_4 is chosen to achieve a specified ARL_0.

SSNEWMA first applies the previous MEWMA scheme to $\boldsymbol{q}_t$ and then, as in Zou et al. (2008), it applies a kernel smoothing to $\boldsymbol{z}_t$:

$$\boldsymbol{f}_t = \boldsymbol{S}\boldsymbol{z}_t ,$$

where $\boldsymbol{S} = (\boldsymbol{S}_n(x_1), \boldsymbol{S}_n(x_2), \ldots, \boldsymbol{S}_n(x_n))'$, $\boldsymbol{S}_n(x_i) = (S_{n1}(x_i), S_{n2}(x_i), \ldots, S_{nn}(x_i))'$, is the kernel smoothing matrix, whose generic element $S_{ni}(x)$ is defined as follows:

$$\begin{aligned}
S_{ni}(x) =& \frac{U_{ni}(x)}{\sum_{j=1}^{n} U_{nj}(x)} \\
U_{nj}(x) =& K_h(x_j - x)[m_{n2}(x) - (x_j - x)m_{n1}(x)] \\
m_{nl}(x) =& \frac{1}{n}\sum_{j=1}^{n} (x_j - x)^l K_h(x_j - x) , \quad l = 1, 2 ,
\end{aligned}$$

where $K_h(u) = \frac{1}{h}K(\frac{u}{h})$ and $K(\cdot)$ is a symmetric probability density function and h a bandwidth. Finally this chart computes the usual MEWMA statistic on $\boldsymbol{f}_t$:

$$\text{SSNEWMA}_t = \frac{2-\lambda}{\lambda}\boldsymbol{f}_t'\boldsymbol{V}\boldsymbol{f}_t\,, \tag{2.9}$$

where $\boldsymbol{V} = \boldsymbol{S} + \boldsymbol{S}' - \boldsymbol{S}'\boldsymbol{S}$. $\frac{2-\lambda}{\lambda}\boldsymbol{V}$ is the inverse of the variance/covariance matrix of $\boldsymbol{f}_t$. The chart signals if $\text{SSNEWMA}_t > h_5$, where h_5 is chosen to achieve a specified ARL_0.

2.2.1 An algorithm to estimate the control limit h

For estimating the control limit h for each one of the control charts, we follow an algorithm introduced in Capizzi & Masarotto (2008), based on imposing that the in-control average run length, $\text{E}_{\text{IC}}(\text{RL})$, is equal to a certain (large) value ARL_0. We give a sketch of this algorithm in the following steps:

1. let h_1 be an initial thought for h; let A, s_0 (the burnin), s (the number of iterations) and α be suitable constants;

2. for $i = 1, \dots, s_0 + s - 1$ repeat:

   (a) simulate a single run length RL_i^* using the control limit h_i;

   (b) update h using the recursive formula

   $$h_{i+1} = \max(0, h_i + Ai^{-\alpha}q_i)\,,$$

   where $q_i = \frac{\text{RL}_i^* - \text{ARL}_0}{\text{ARL}_0}$ and the max operator is used to ensure that the h estimates are positive at each step;

3. discard the first s_0 values and estimate the control limit h using

$$\hat{h}_s = \frac{1}{s - s_0}\sum_{j=s_0+1}^{s_0+s} h_i\,.$$

In order to avoid very long execution times, due to extremely large run lengths, a truncation can be introduced: it is convenient to arrest the run length simulation at the value MRL_i such that $h_{i+1} - h_i \leq \xi$ (which means $\mathrm{MRL}_i \leq \mathrm{ARL}_0(1 + \xi A^{-1} i^{\alpha})$), where ξ is a constant for which values 2 or 3 are suggested.

In this way, we keep the sequence h_i from wandering too much and also avoid useless long simulation runs.

On the basis of this algorithm, in literature also another algorithm is used, based on imposing that the probability of false alarm before some specified value N_0 is equal to a desired (small) value p_0, but in this work we used only the first method to estimate h.

2.3 Simulation Study and Performance Comparison

In this section we investigate the monitoring performance of the five proposed profile monitoring schemes through ARL comparisons in a simulation study we are going to describe; for simulations we are going to describe we used a code written both in `R` (R Development Core Team, 2010) and in `C` languages: this makes the code much faster that it would be using only `R`, indeed `C` enhances the part of the simulation of the run length above all.

Actually, the first control chart, SSMEWMA, is already existing, since it is simply the Multivariate EWMA (MEWMA) arranged to profile data; thus it is more correct to say that performances of the other four control schemes are compared to performances of the already existing SSMEWMA charting scheme.

Consistently with recent literature about profile monitoring (Zou et al., 2008; Capizzi & Masarotto, 2011), for our simulation study we consider the IC model

$$y_{i,t} = \mu_0(x_i) + \epsilon_{i,t}\,, \quad \mu_0(x) = 1 - \exp(-x)\,, \quad i = 1, \ldots, n\,,\ t = 1, 2, \ldots \tag{2.10}$$

and three OC models

$$y_{i,t} = \mu_j(x_i) + \epsilon_{i,t}\,, \quad j = 1, 2, 3\,, \tag{2.11}$$

where

$$\mu_1(x;\beta_1,\beta_2) = 1 - \beta_1 \exp(-\beta_2 x) \tag{2.12}$$

$$\mu_2(x;\beta_1,\beta_2) = 1 - \exp(-x) + \beta_1 \cos[\beta_2 \pi (x - 0.5)] \tag{2.13}$$

$$\mu_3(x;\beta_1,\beta_2) = 1 - \exp\left\{-x - \beta_1 \left[\max\left(\frac{x-\beta_2}{1-\beta_2}\right)\right]^2\right\}\,. \tag{2.14}$$

Model	Label	β_1	β_2	ν
1	A	1.00	1.20	0.192
1	B	1.10	1.00	0.372
1	C	1.00	1.50	0.428
1	D	1.20	1.00	0.744
1	E	1.30	1.00	1.116
1	F	1.60	1.00	2.232
2	A	0.10	2.00	0.400
2	B	0.10	3.00	0.400
2	C	0.20	4.00	0.800
2	D	0.20	5.00	0.800
2	E	0.30	3.00	1.200
2	F	0.60	3.00	2.400
3	A	2.00	0.90	0.331
3	B	4.00	0.90	0.444
3	C	2.00	0.75	0.545
3	D	4.00	0.75	0.732
3	E	5.00	0.50	1.195
3	F	10.00	0.10	2.205

Table 2.1: Values of the parameters of the 18 OC models used in simulations.

For every OC model we consider six different combinations of the two parameters, taken in Capizzi & Masarotto (2011) and reported in table 2.1, in order to obtain $6 \times 3 = 18$ different OC models, which are illustrated in figg. 2.1 (model 2.12), 2.2 (model 2.13) and 2.3 (model 2.14). As you can see in the figures, model 2.12 represents a series of shifted models which are, in a sense, parallel to the IC model, model 2.13 provides a kind of wave-shaped shift and model 2.14 represents a local shift (from a certain point on the OC curve begins to differ from the IC one). In table 2.1 we also reported for each model the values of the global out-of-control parameter $\nu = \sqrt{\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}}$, which in this case is simply $\nu = \sqrt{\boldsymbol{\delta}'\boldsymbol{\delta}}$, since $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n = \boldsymbol{I}_n$.

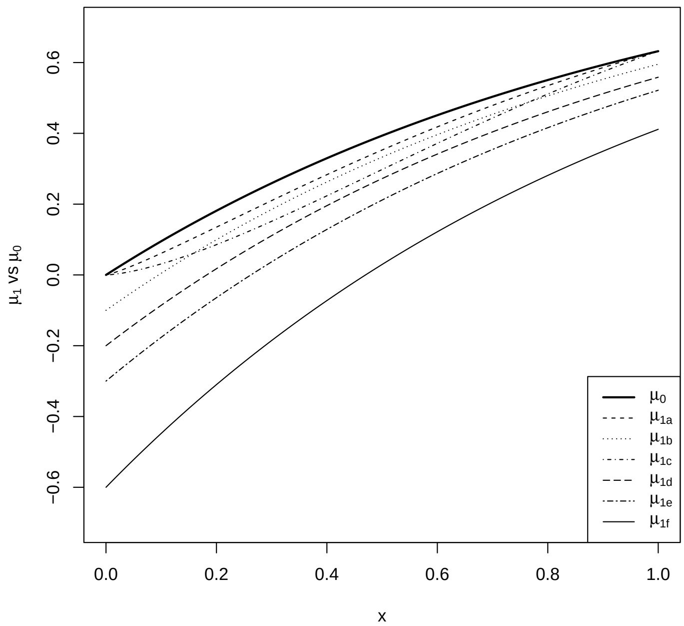

Figure 2.1: Comparison between IC and OC1 models.

Notice that OC1 model reduces to IC model when $\beta_1 = \beta_2 = 1$ and the same happens for OC2 and OC3 models when $\beta_1 = 0$. For each of these 18 OC models $\sigma^2 = 1$ and 5 possible values of the change-point τ are considered: 51, 101, 151, 201 and 301. Therefore we apply the presented charting schemes to $3 \times 6 \times 5 = 90$ different OC scenarios. Moreover, as in Zou et al. (2008), we restrict our study to the equally spaced design points for the explanatory

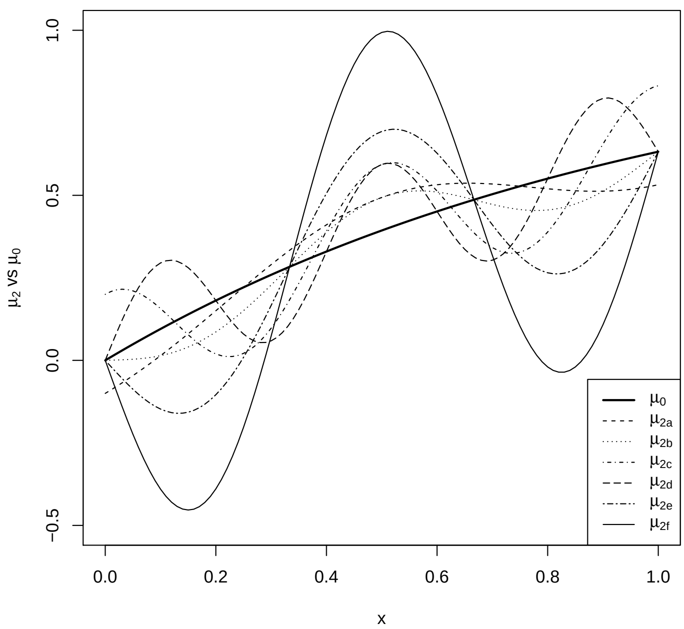

Figure 2.2: Comparison between IC and OC2 models.

variable $\boldsymbol{x}$:

$$x_i = \frac{i - 0.5}{n}, \quad i = 1, \dots, n,$$

with $n = 32$.

We consider also different choices for the parameters of the charting schemes in order to investigate their role:

- for SSMEWMA chart: 3 different values for λ (0.025, 0.050 and 0.200);
- for SSWFMEWMA chart: 3 different values for λ (0.025, 0.050 and 0.200), 2 different values for k (1.5 and 3.5, which correspond to high quantiles, more or less 0.8664 and 0.9995 quantiles, respectively, of the IC distribution of $\sqrt{\frac{2-\lambda}{\lambda}}|w_{i,t}|$, $i = 1, \dots, n$, $t = 2, 3, \dots$, which is approximately standard normal) and 2 different wavelet transforms: Least Asymmetric 10 (La10) and Haar transforms (see Nason (2008) and Doroslovački (1998) for further details about their definition);

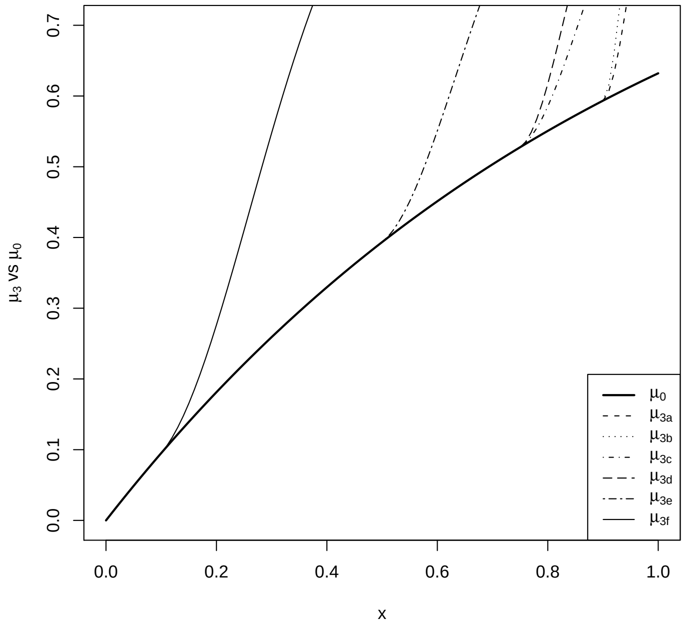

Figure 2.3: Comparison between IC and OC3 models.

- for SSCUSCORE chart: only one combination, since it does not depend on λ, k or any wavelet transformation;
- for SSWFCUSCORE chart: 3 values for k, chosen depending on the empirical IC distribution of the CUSCORE distribution (we made a small pilot simulation in order to get it: 0, 2.2 and 4, which correspond to high quantiles, more or less 0, 0.8 and 0.995 quantiles respectively, of the empirical IC distribution of $\max\left(-C^L_{i,t}, C^U_{i,t}\right)$) and 2 different wavelet transforms (La10 and Haar); note that, differently from SSWFMEWMA, where if $k = 0$ it reduces to SSWFMEWMA, SSWFCUSCORE with $k = 0$ is different from SSCUSCORE;
- for SSNEWMA chart: 3 different values for λ (0.025, 0.050 and 0.200).

Some charting schemes have also other minor parameters, which we have taken to be equal to the values suggested in literature. In SSCUSCORE and

SSWFCUSCORE schemes we consider $m = 0.25$ and $c = 0.50$, as suggested in Capizzi & Masarotto (2010b); in SSNEWMA scheme we use for simplicity, as in Zou et al. (2008), Epanechnikov's kernel, which is

$$K_E(u) = \frac{3}{4}(1 - u^2)\mathbb{I}_{[-1,1]}(u)\,,$$

with bandwidth computed as in Zou et al. (2008):

$$h = gn^{-\frac{1}{5}} \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^{\frac{1}{2}}\,,$$

where g can empirically be any value in the interval $[1.0, 2.0]$; we estimate g such that we get just the quantity of smoothing imposed (in this case we impose 6 equivalent degrees of freedom).

We therefore obtain 3 SSMEWMA, $3 \times 2 \times 2 = 12$ SSWFMEWMA, 1 SSCUSCORE, $3 \times 2 = 6$ SSWFCUSCORE and 3 SSNEWMA schemes combinations of charts, 25 different combinations all together. We estimate each of the control limits of these 25 combinations of charting schemes (these estimates are reported in the second column of table 2.2) always imposing $\text{ARL}_0 = 500$ and using an appropriate algorithm (with 10000 iterations) used also in Capizzi & Masarotto (2008). The first column of table 2.2 reports the names of the 25 control schemes, where one digit in the names of the scheme (SSMEWMA and SSNEWMA) refers to $\lambda = 0.025, 0.05, 0.2$, two digits (SSWFCUSCORE) refer to La10 and Haar wavelet transform (see Nason (2008) for their definition) and $k = 0, 2.2, 4$ respectively, and three digits (SSWFMEWMA) refer to $\lambda = 0.025, 0.05, 0.2$, La10 and Haar wavelet transform and $k = 1.5, 3.5$ respectively.

In order to estimate in a quite robust way the OC ARLs, for each of the 90 OC scenarios, for each of the 25 charting schemes, 50000 observations of the Run Length (RL) are simulated.

Chart	h	RMI	RMI_1	RMI_2	RMI_3	RMI_a	RMI_b	RMI_c
SSMEWMA1	52.02	1.00	1.21	0.50	1.30	0.91	0.87	1.49
SSMEWMA2	55.03	1.36	1.56	0.73	1.79	1.55	0.95	1.01
SSMEWMA3	58.96	4.05	4.35	2.93	4.89	4.22	6.77	0.66
SSWFMEWMA111	38.65	0.99	1.19	0.50	1.29	0.88	0.89	1.56
SSWFMEWMA112	12.89	1.01	0.86	0.53	1.63	1.09	0.61	1.08
SSWFMEWMA121	38.77	1.02	1.23	0.52	1.32	0.92	0.90	1.56
SSWFMEWMA122	12.86	1.15	0.85	0.98	1.62	1.18	0.86	1.31
SSWFMEWMA211	41.80	1.42	1.64	0.77	1.86	1.61	1.02	1.07
SSWFMEWMA212	14.06	1.42	1.12	0.84	2.29	1.82	0.57	0.65
SSWFMEWMA221	41.81	1.42	1.63	0.78	1.86	1.62	1.00	1.06
SSWFMEWMA222	14.08	1.65	1.14	1.46	2.35	2.02	0.99	0.85
SSWFMEWMA311	45.38	4.05	4.37	2.93	4.85	4.17	6.90	0.72
SSWFMEWMA312	15.68	3.81	3.14	2.92	5.36	4.47	4.75	0.19
SSWFMEWMA321	45.50	4.20	4.53	3.04	5.02	4.33	7.14	0.72
SSWFMEWMA322	15.65	4.20	3.10	4.17	5.33	4.52	6.66	0.47
SSCUSCORE	62.91	1.41	1.60	0.86	1.77	0.69	2.33	3.37
SSWFCUSCORE11	62.93	2.21	2.83	1.39	2.40	0.78	3.45	6.67
SSWFCUSCORE12	11.94	1.92	2.31	1.24	2.22	0.82	3.02	5.22
SSWFCUSCORE13	2.21	1.28	1.21	0.82	1.80	0.78	1.77	2.77
SSWFCUSCORE21	63.02	2.13	2.87	1.06	2.45	0.78	3.28	6.36
SSWFCUSCORE22	12.05	1.91	2.35	1.14	2.23	0.86	2.96	5.03
SSWFCUSCORE23	2.19	1.40	1.20	1.22	1.79	0.83	2.03	3.06
SSNEWMA1	13.05	0.21	0.19	0.28	0.16	0.10	0.14	0.72
SSNEWMA2	14.65	0.32	0.25	0.53	0.20	0.39	0.00	0.39
SSNEWMA3	17.02	1.59	1.13	2.33	1.29	2.15	0.91	0.00

Table 2.2: Estimated control limits and RMI index for the 25 control charts used in simulations.

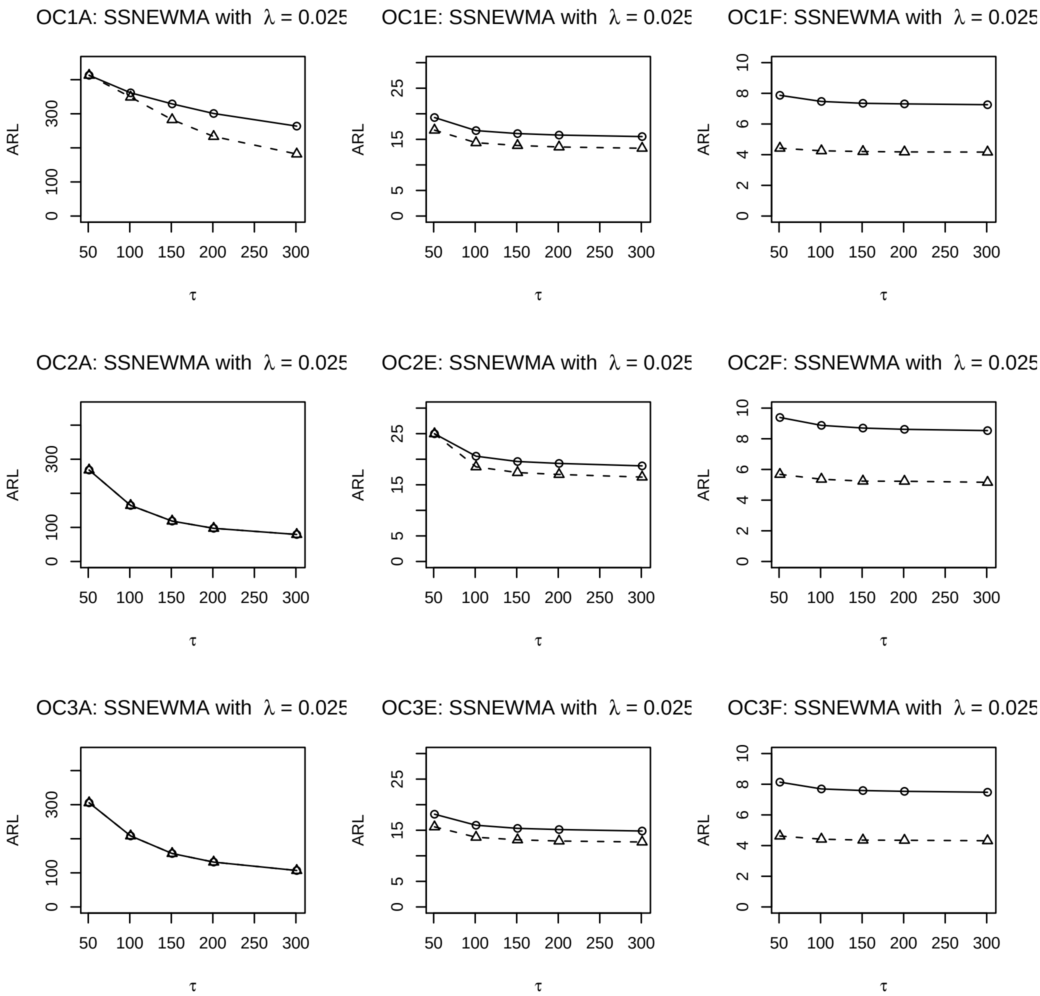

Figure 2.4: Study of performaces of SSNEWMA chart with $\lambda = 0.025$, by some scenarios and values of τ, with respect to other charts.

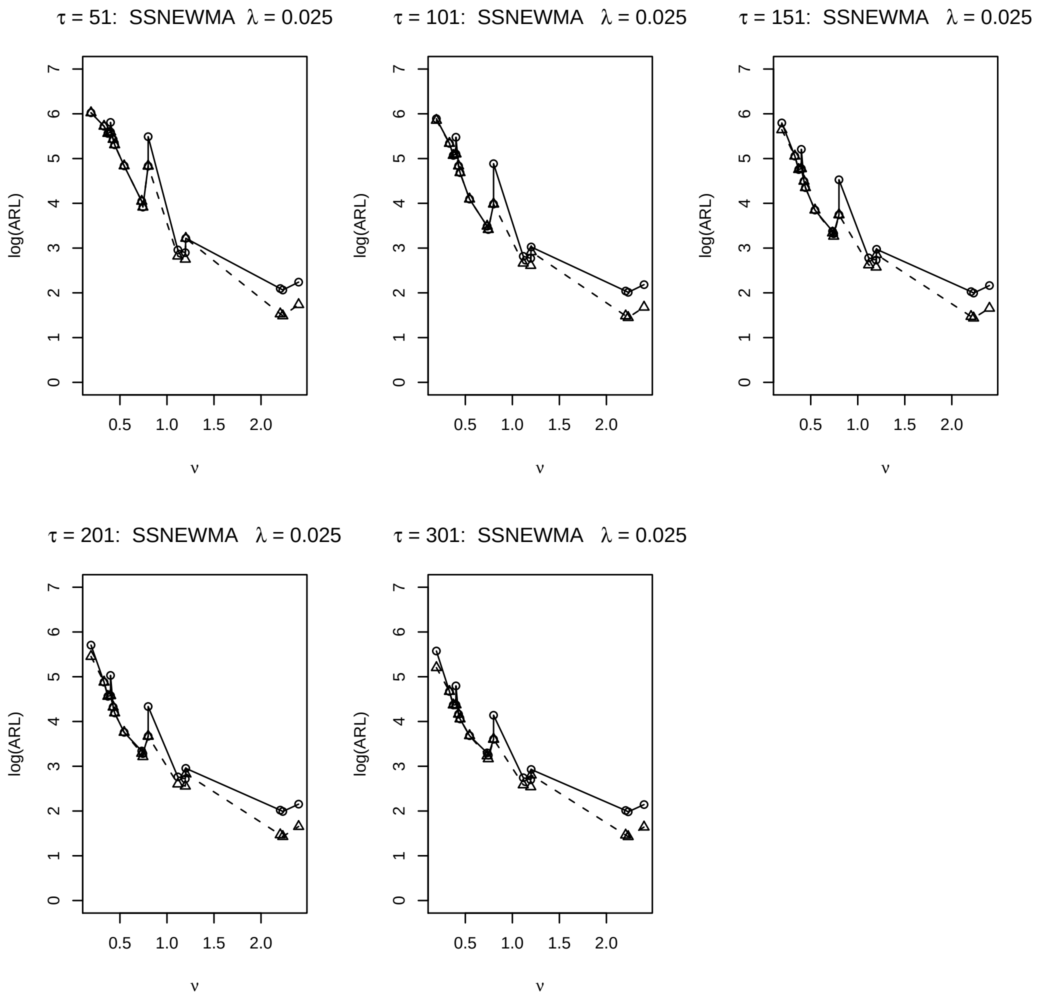

Figure 2.5: Study of performaces of SSNEWMA chart with $\lambda = 0.025$, by values of τ and values of ν, with respect to other charts.

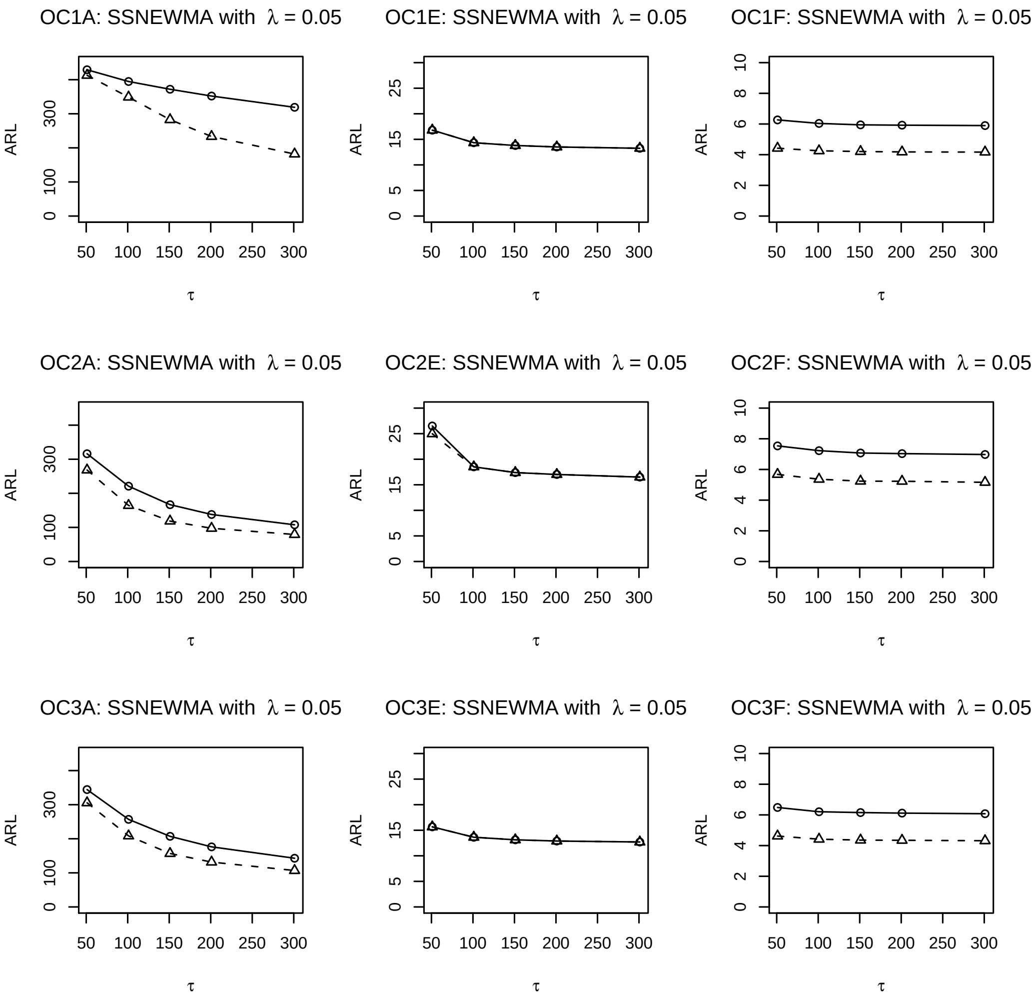

Figure 2.6: Study of performaces of SSNEWMA chart with $\lambda = 0.05$, by some scenarios and values of τ, with respect to other charts.

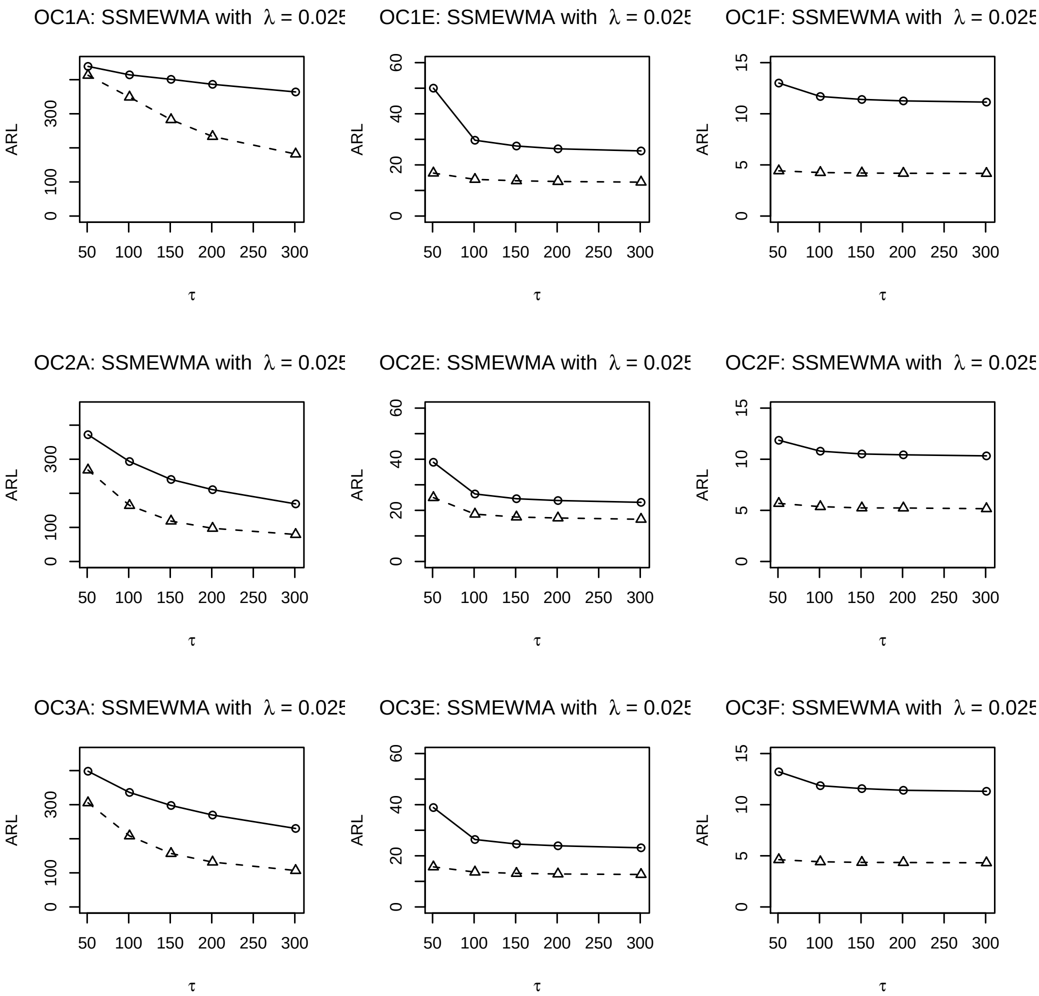

Figure 2.7: Study of performaces of SSMEWMA chart with $\lambda = 0.025$, by some scenarios and values of τ, with respect to other charts.

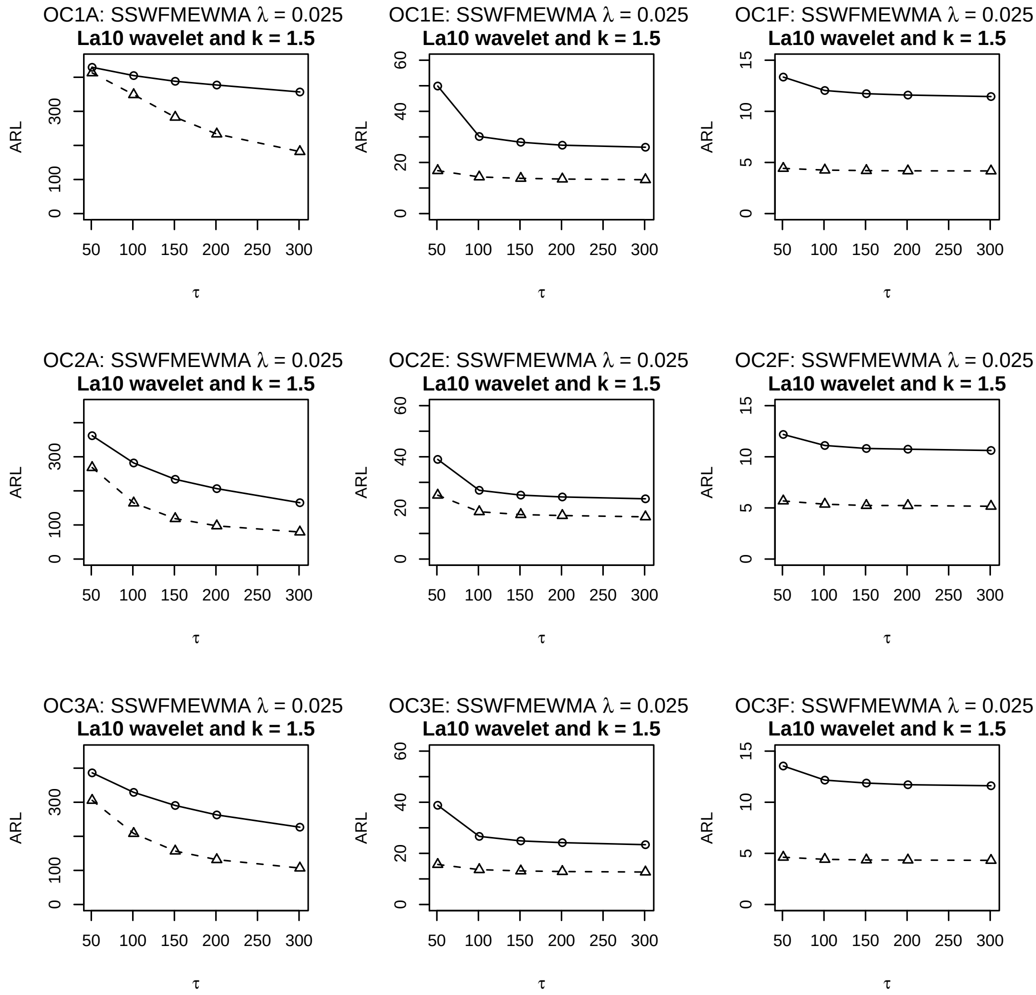

Figure 2.8: Study of performaces of SSWFMEWMA chart with $\lambda = 0.025$, La10 wavelet and $k = 1.5$, by some scenarios and values of τ, with respect to other charts.

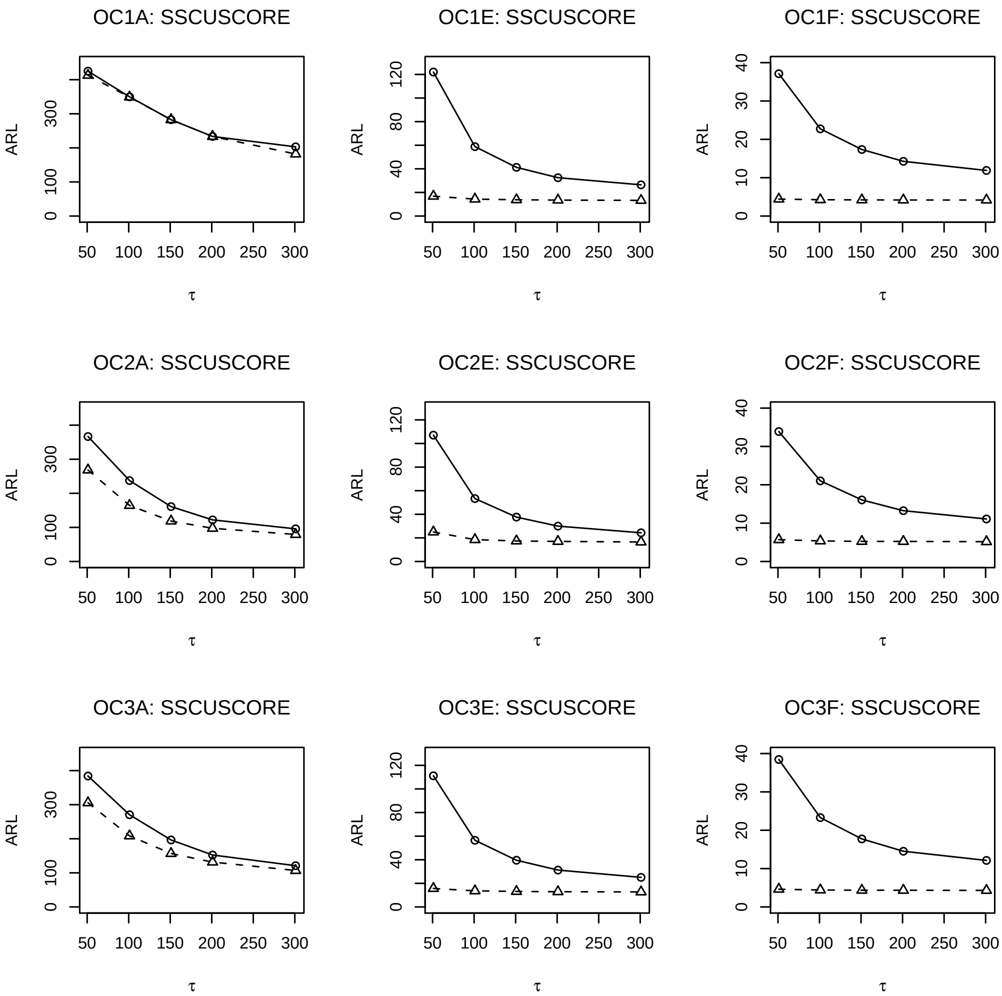

Figure 2.9: Study of performaces of SSCUSCORE chart, by some scenarios and values of τ, with respect to other charts.

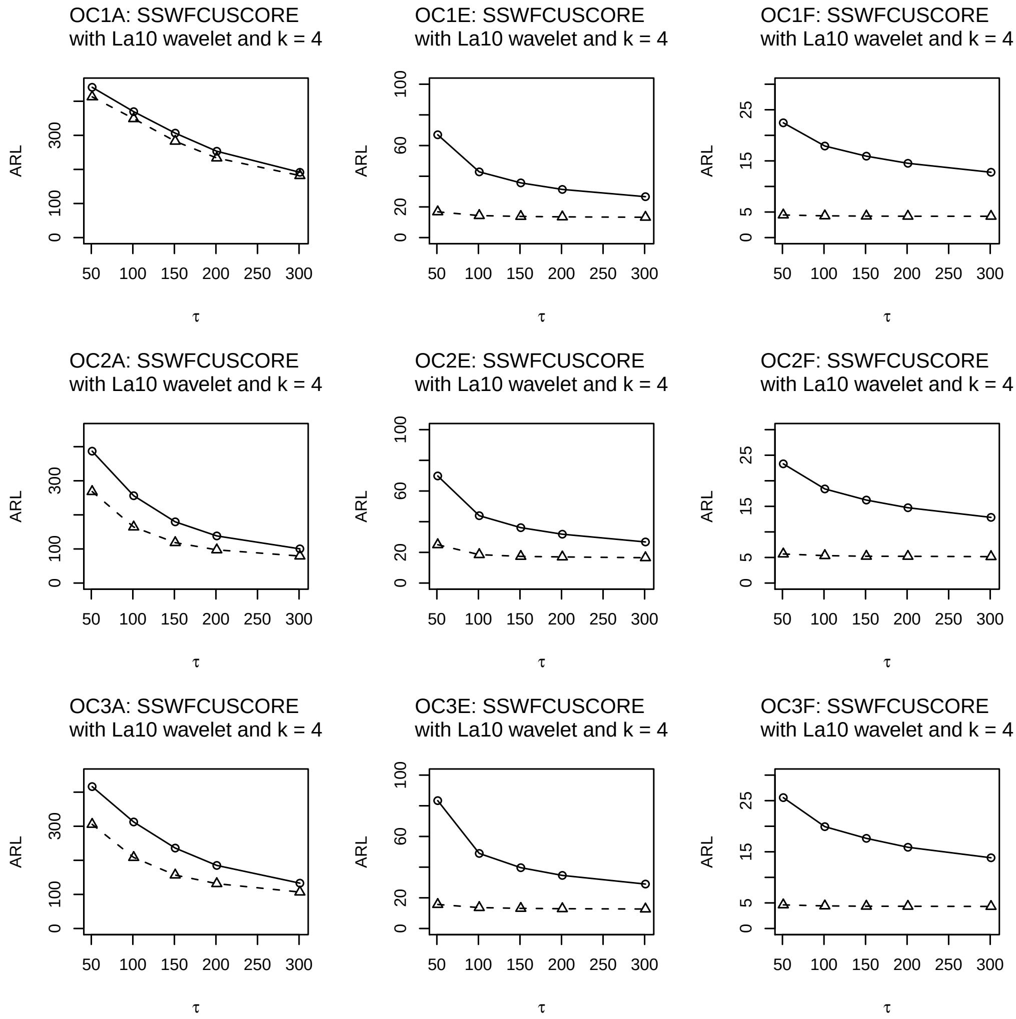

Figure 2.10: Study of performaces of SSWFCUSCORE chart with La10 wavelet and $k = 1.5$, by some scenarios and values of τ, with respect to other charts.

A quite good method for evaluating control charts and summarizing their performances is to calculate their Relative Mean Index (RMI) index introduced in Han & Tsung (2006), which is a summary performance measure defined as

$$\mathrm{RMI} = \frac{1}{N}\sum_{r=1}^{N}\frac{\mathrm{ARL}_{\delta_r} - \mathrm{MARL}_{\delta_r}}{\mathrm{MARL}_{\delta_r}}\,, \qquad (2.15)$$

where N is the total number of considered shifts, ARL_{δ_r} the ARL of a chart for detecting a shift δ_r and MARL_{δ_r} is the smallest ARL among the ARL values of the compared charts. Hence, RMI measures the average relative efficiency for a range of shift sizes, and a control chart with a smaller RMI is considered better in its overall performance.

Columns 3-9 of table 2.2 report the RMI indices of the 25 investigated control charts: in the third column the RMI index is global and includes all 90 scenarios ($N = 90$); in the next three columns we reported the RMI indices computed on OC1 (RMI_1, $N_1 = 30$), OC2 (RMI_2, $N_2 = 30$) and OC3 (RMI_3, $N_3 = 30$) scenarios, respectively; the last three columns contain the RMI indices computed on scenarios with small shifts from the IC model (RMI_a, $\nu < 1$: see labels A, B, C and D of table 2.1) ($N_a = 60$), medium shifts (RMI_b, $1 \leq \nu < 2$: label E) ($N_b = 15$) and large shifts (RMI_c, $\nu \geq 2$: label F) ($N_c = 15$). There are more scenarios with $\nu < 1$, because in this work we are interested mainly in small deviations from the IC model.

To summarize clearly the results of the simulations, we cannot represent all the perfomances of all 25 charts together, therefore we compared each chart with all other charts together by representing in each graph the ARLs of the considered chart (by values of τ for a certain scenario or by values of ν for a certain τ) compared with the smallest ARL among the other 24 charts. In the graphs highlighting the role of τ we report only the ARLs in models OC1A, OC1E, OC1F, OC2A, OC2E, OC2F, OC3A, OC3E, OC3F, which are, for each of the three types of OC model, the scenarios with the smallest, a medium and the highest values of the OC parameter ν, whereas in the graphs highlighting the role of ν we have reported the ARLs (actually, in this case, the logarithm of the ARLs, to be able to better read the graphs) in all 18 OC models. In both types of graphs the more the line of the considered chart (the solid one) is near the line of the best charts (the dashed one), the better is the performance of the considered chart.

In the figures of this paragraph there are only some of the results: in figg. 2.4 and 2.5 one can see the performances of the SSNEWMA charts with $\lambda = 0.025$ in terms of ARL highlighting the role of τ in the 9 models of greatest interest and the role of ν in all 18 models. This chart seems to behave in general in a better way than the other ones, especially in case of not so large shifts. Even with a value of λ slightly higher (0.05) the SSNEWMA chart seems to behave in a good way (see fig. 2.6); in this case the performance of SSNEWMA chart is better in case of medium and large shifts, whereas it gets a little worse in case of small shifts. In figg. 2.7, 2.8, 2.9 and 2.10 you can find the performances (only graphs highlighting the role of τ) of the best (in terms of global RMI index) SSMEWMA chart ($\lambda = 0.025$), the best SSWFMEWMA chart ($\lambda = 0.025$, La10 wavelet and $k = 1.5$), the SSCUSCORE chart and the best SSWFCUSCORE chart (La10 wavelet and $k = 4$), which seem to have often or always worse performances of a SSNEWMA chart with a good choice of the smoothing parameter λ.

Have a look at table 2.2, where RMI indices are reported; in table 2.3 there are seven rankings (global, OC1, OC2, OC3, small shifts, medium shifts, large shifts) made by ordering the 25 control charts investigated by RMI index (these rankings are obtained simply by ranking charts by RMI indices reported in table 2.2): this helps us to try to sum up the results of this simulation study into the following points:

- there are some slight differences of performances in the three different groups of scenarios, but in general the control schemes behave more or less in the same way with respect to the model;
- SSNEWMA, with a good choice of λ, seems to have almost always the best performance in detecting a shift from the expected model;
- for SSNEWMA, SSMEWMA and SSWFMEWMA λ has to be chosen

	RRMI	$RRMI_1$	$RRMI_2$	$RRMI_3$	$RRMI_a$	$RRMI_b$	$RRMI_c$
1	SSNEWMA1	SSNEWMA1	SSNEWMA1	SSNEWMA1	SSNEWMA1	SSNEWMA2	SSNEWMA3
2	SSNEWMA2	SSNEWMA2	SSWFMEWMA111	SSNEWMA2	SSNEWMA2	SSNEWMA1	SSWFMEWMA312
3	SSWFMEWMA111	SSWFMEWMA122	SSMEWMA1	SSNEWMA3	SSCUSCORE	SSWFMEWMA212	SSNEWMA2
4	SSMEWMA1	SSWFMEWMA112	SSWFMEWMA121	SSWFMEWMA111	SSWFCUSCORE13	SSWFMEWMA112	SSWFMEWMA322
5	SSWFMEWMA112	SSWFMEWMA212	SSNEWMA2	SSMEWMA1	SSWFCUSCORE11	SSWFMEWMA122	SSWFMEWMA212
6	SSWFMEWMA121	SSNEWMA3	SSWFMEWMA112	SSWFMEWMA121	SSWFCUSCORE21	SSMEWMA1	SSMEWMA3
7	SSWFMEWMA122	SSWFMEWMA222	SSMEWMA2	SSWFMEWMA122	SSWFCUSCORE12	SSWFMEWMA111	SSNEWMA1
8	SSWFCUSCORE13	SSWFMEWMA111	SSWFMEWMA211	SSWFMEWMA112	SSWFCUSCORE23	SSWFMEWMA121	SSWFMEWMA311
9	SSMEWMA2	SSWFCUSCORE23	SSWFMEWMA221	SSCUSCORE	SSWFCUSCORE22	SSNEWMA3	SSWFMEWMA321
10	SSWFCUSCORE23	SSMEWMA1	SSWFCUSCORE13	SSMEWMA2	SSWFMEWMA111	SSMEWMA2	SSWFMEWMA222
11	SSCUSCORE	SSWFCUSCORE13	SSWFMEWMA212	SSWFCUSCORE23	SSMEWMA1	SSWFMEWMA222	SSMEWMA2
12	SSWFMEWMA212	SSWFMEWMA121	SSCUSCORE	SSWFCUSCORE13	SSWFMEWMA121	SSWFMEWMA221	SSWFMEWMA221
13	SSWFMEWMA221	SSMEWMA2	SSWFMEWMA122	SSWFMEWMA221	SSWFMEWMA112	SSWFMEWMA211	SSWFMEWMA211
14	SSWFMEWMA211	SSCUSCORE	SSWFCUSCORE21	SSWFMEWMA211	SSWFMEWMA122	SSWFCUSCORE13	SSWFMEWMA112
15	SSNEWMA3	SSWFMEWMA221	SSWFCUSCORE22	SSWFCUSCORE12	SSMEWMA2	SSWFCUSCORE23	SSWFMEWMA122
16	SSWFMEWMA222	SSWFMEWMA211	SSWFCUSCORE23	SSWFCUSCORE22	SSWFMEWMA211	SSCUSCORE	SSMEWMA1
17	SSWFCUSCORE22	SSWFCUSCORE12	SSWFCUSCORE12	SSWFMEWMA212	SSWFMEWMA221	SSWFCUSCORE22	SSWFMEWMA111
18	SSWFCUSCORE12	SSWFCUSCORE22	SSWFCUSCORE11	SSWFMEWMA222	SSWFMEWMA212	SSWFCUSCORE12	SSWFMEWMA121
19	SSWFCUSCORE21	SSWFCUSCORE11	SSWFMEWMA222	SSWFCUSCORE11	SSWFMEWMA222	SSWFCUSCORE21	SSWFCUSCORE13
20	SSWFCUSCORE11	SSWFCUSCORE21	SSNEWMA3	SSWFCUSCORE21	SSNEWMA3	SSWFCUSCORE11	SSWFCUSCORE23
21	SSWFMEWMA312	SSWFMEWMA322	SSWFMEWMA312	SSWFMEWMA311	SSWFMEWMA311	SSWFMEWMA312	SSCUSCORE
22	SSWFMEWMA311	SSWFMEWMA312	SSMEWMA3	SSMEWMA3	SSMEWMA3	SSWFMEWMA322	SSWFCUSCORE22
23	SSMEWMA3	SSMEWMA3	SSWFMEWMA311	SSWFMEWMA321	SSWFMEWMA321	SSMEWMA3	SSWFCUSCORE12
24	SSWFMEWMA321	SSWFMEWMA311	SSWFMEWMA321	SSWFMEWMA322	SSWFMEWMA312	SSWFMEWMA311	SSWFCUSCORE21
25	SSWFMEWMA322	SSWFMEWMA321	SSWFMEWMA322	SSWFMEWMA312	SSWFMEWMA322	SSWFMEWMA321	SSWFCUSCORE11

Table 2.3: Ranking of the charts by RMI index.

in a proper way and in these cases, where all shifts are quite small, a small value for λ (0.05 or, even better, 0.025) gives better performances of the charts;

- even with a large shift, a small value of λ is not so bad as a large value of λ in presence of a small shift, that is in general a small λ is a more conservative choice;

- type of wavelet transform, even if we tried only two types (La10 and Haar wavelets), seems not to have a strong influence on performances of control schemes;

- it seems more difficult to understand the role of the quantity of thresholding in SSWFMEWMA: in general we could say that a certain quantity of smoothing, but not too much, seems useful to improve the performance of SSMEWMA (which is, indeed, a SSWFMEWMA without any thresholding);

- a high value of the thresholding parameter k often makes SSWFCUSCORE better than SSCUSCORE, but in general both these charts have worse performances than other schemes (even if in case of small shifts they behave quite good);

- the role of the change point τ is rather clear: the more it is forward over time, the better are the performances of these charts, because self-starting procedures have in this case more IC observations at their disposal;

- in general control charts can detect before a larger shift from the IC model, but when we deal with values of ν which are very close, the speed of detecting an OC signal depends also on the type of the OC model.

2.4 Other Attempts to Improve This Work

2.4.1 Adaptive λ

Another direction of improvement of this work could be to propose an adaptive version of SSMEWMA, SSWFMEWMA and SSNEWMA, by using a smoothing parameter λ which changes and adapts to the magnitude of the shift as the observations are collected.

In particular, since in general (and also, as you have seen before, in our framework) the larger the value of λ is, the more able the chart to recognize larger shifts (with a larger OC parameter ν) is, the adaptive λ should become smaller at a certain time t if the chart realizes that in that time a small deviation from the IC situation or no deviation at all occurs, whereas it should become larger if the chart realizes that in that time there is a large deviation from the IC situation.

The first example of adaptive chart was introduced by Sparks (2000), which is essentially an adaptive version of a CUSUM chart. In Capizzi & Masarotto (2003), instead, we can see a first proposal of adaptive EWMA control chart, the so called AEWMA (Adaptive Exponentially Weighted Moving Average), more useful to our objectives. Mahmoud & Zahran (2010) introduced a multivariate version of this AEWMA chart, the so-called MAEWMA chart. In Capizzi & Masarotto (2010a), instead, you can find a first proposal of adaptive and multivariate EWMA chart applied to profile data, in particular to the three parameters of a simple linear regression model (the regression coefficients β_0 and β_1, and the variance of the error term σ^2).

Since SSNEWMA chart seems to be the most performing among the five control charts we proposed in this work, we try to enhance only its performance by making λ adaptive. The procedure we used to do this is described in the following lines.

SSNEWMA chart is based on the usual MEWMA scheme applied to $\boldsymbol{q}_t$

$$\boldsymbol{z}_t = (1-\lambda)\boldsymbol{z}_{t-1} + \lambda \boldsymbol{q}_t \,,$$

with $\boldsymbol{z}_0 = \mathbf{0}_n$ and $\lambda \in (0,1)$; it can be rewritten in the following way:

$$\boldsymbol{z}_t = \boldsymbol{z}_{t-1} + \lambda(\boldsymbol{q}_t - \boldsymbol{z}_{t-1})\,.$$

SSNAEWMA , instead, is based on a MAEWMA (Multivariate Adaptive EWMA) scheme which, differently from the MEWMA, substitutes λ with $w(e_t)$, a weight depending on e_t the euclidean distance between $\boldsymbol{q}_t$ and $\boldsymbol{z}_{t-1}$ ($||\boldsymbol{q}_t - \boldsymbol{z}_{t-1}||$), which means making λ depending on the magnitude of the shift the chart is analyzing:

$$\boldsymbol{z}_t = \boldsymbol{z}_{t-1} + w(e_t)(\boldsymbol{q}_t - \boldsymbol{z}_{t-1})\,,$$

where

$$w(e_t) = \frac{\phi(e_t)}{e_t}\,, \quad e_t = ||\boldsymbol{q}_t - \boldsymbol{z}_{t-1}|| = \sqrt{\sum_{i=1}^{n} (q_{i,t} - z_{i,t-1})^2}\,.$$

As explained in Capizzi & Masarotto (2003), the score function $\phi(\cdot)$ is thought to make the chart behave as a smooth combination of an EWMA chart, more able to detect small shifts, and a Shewhart chart, more able to detect large shifts. To do this, $\phi(\cdot)$ has to satisfy:

1. $\phi(x) = -\phi(-x)$ (odd function);
2. $\frac{\phi(x)}{x} \doteq \lambda$ if $|x|$ is small (i.e., behaving like EWMA);
3. $\frac{\phi(x)}{x} \doteq 1$ if $|x|$ is large (i.e., behaving like Shewhart).

In their article Capizzi & Masarotto (2003) discussed three score functions that satisfy these conditions, but in the present article, as in Mahmoud & Zahran (2010), we will focus on their first score function

$$\phi(e_t) = \begin{cases} e_t - (1-\lambda)k & \text{if } e_t < -k \\ \lambda e_t & \text{if } -k \le e_t \le k \\ e_t - (1-\lambda)k & \text{if } e_t > k\,, \end{cases}$$

where $\lambda \in (0,1)$ and $k > 0$. Since we are in a multivariate framework and e_t, being an euclidean distance, cannot be negative, the particular definition of $\phi(\cdot)$ in our case is

$$\phi(e_t) = \begin{cases} \lambda e_t & \text{if } 0 \leq e_t \leq k \\ e_t - (1-\lambda)k & \text{if } e_t > k\,. \end{cases} \tag{2.16}$$

Then, as in SSNEWMA, SSNAEWMA applies a kernel smoothing to this alternative version of $\boldsymbol{z}_t$:

$$\boldsymbol{f}_t = \boldsymbol{S}\boldsymbol{z}_t\,,$$

where $\boldsymbol{S}$ is the kernel smoothing matrix, defined as in SSNEWMA (see page 18). Finally this chart computes the usual MEWMA statistic on $\boldsymbol{f}_t$:

$$\text{SSNAEWMA}_t = \frac{2-\lambda}{\lambda}\boldsymbol{f}_t'\boldsymbol{V}\boldsymbol{f}_t\,, \tag{2.17}$$

where $\boldsymbol{V}$ is defined as in SSNEWMA.
The chart signals if $\text{SSNAEWMA}_t > h_6$, where h_6 is chosen to achieve a specified ARL_0.

A simulation study has been conducted to compare performances of SSNAEWMA with those of SSNEWMA and the same scenarios of section 2.3 have been used: 1 IC and 18 OC models (6 for each of the three types of deviations from the IC model considered) combined with 5 values (51,101,151,201 and 301) of the change point τ, getting $1+18\times 5 = 91$ scenarios all together. We used the same three values of λ already used (0.025, 0.05 and 0.1), getting $91 \times 3 = 273$ different ARLs which can be compared to learn something more about this comparison.

Results can be summarized in the following way:

- in scenarios with an OC deviation of the first type, parallel to the IC model (see fig. 2.1), SSNAEWMA seems to perform always better than SSNEWMA (ARLs are from 2 to 31% lower);

- in scenarios with an OC deviation of the second type, a kind of wave-shaped shift (see fig. 2.2), SSNAEWMA seems to perform always better than SSNEWMA except for scenario 2C, which has a medium value of the OC parameter ν;

- in scenarios with an OC deviation of the third type, representing a local shift (see fig. 2.3), SSNAEWMA seems to perform better than SSNEWMA only when the OC parameter ν is small;

- in general, as we expected, usually when ν is small SSNAEWMA has a larger gain in terms of ARL with respect to SSNEWMA with a large value of λ and vice versa. However, this does not happen always.

Therefore, unless AEWMA (or MAEWMA) is always a guarantee of enhanced performance with respect to EWMA (or MEWMA), that is in univariate or simple multivariate frameworks, the behaviour of SSNAEWMA seems to be still a bit contradictory with profile data and maybe its properties and capabilities should be better investigated. Furthermore, some other tricks may be necessary to use adaptive charts in this nonparametric profile monitoring framework.

2.4.2 Something More About the Correct Number of Degrees of Freedom in SSNEWMA

In case of SSNEWMA chart, we imposed 6 as number of equivalent parameters because we thought this choice to be the usual good compromise between goodness of fit and smoothness which we have always to make when we are dealing with every kind of smoothing parameter.

We would like to try to justify this choice in this work, therefore we made also a brief simulation study on performances of SSNEWMA chart in function of the number of degrees of freedom. In particular, in the simulation study, we chose values 3, 6, 9 and 16 as possible degrees of freedom and for each of this value we have the same 273 ARLs of before; in figures 2.11, 2.12 and 2.11 we find 9 of them (scenarios OC1A with $\tau = 201$, OC2B with $\tau = 151$

and OC3B with $\tau = 51$, and all three scenarios with all three values of λ considered, 0.025, 0.05 and 0.1) as representative examples.

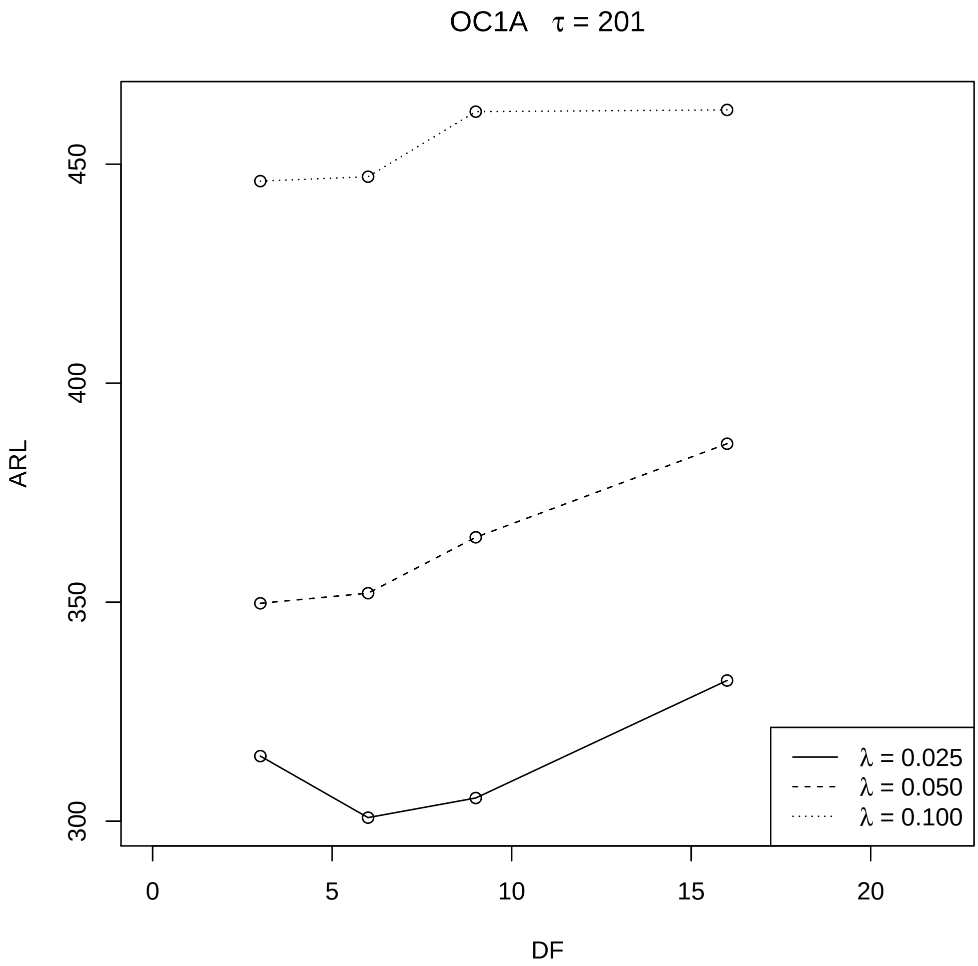

Figure 2.11: Example 1 on degrees of freedom of SSNEWMA.

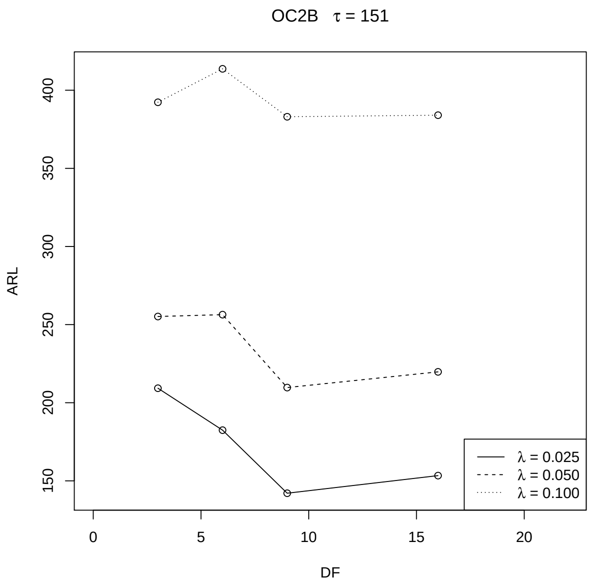

Figure 2.12: Example 2 on degrees of freedom of SSNEWMA.

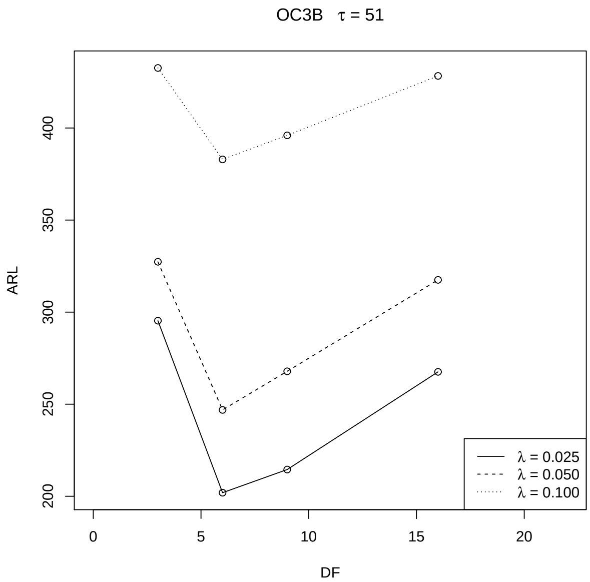

Figure 2.13: Example 3 on degrees of freedom of SSNEWMA.

In the examples in the graphs above, the choice of 6 as the number of degrees of freedom for simulation studies with SSNEWMA chart is a good choice, since ARLs in this case are the smallest ones among ARLs obtained with 3, 6, 9 or 16 degrees of freedom. These are only three examples of all 91 scenarios we tested, but almost always ARLs obtained with 6 degrees of freedom are the smallest ones or, at least, very close to the smallest ones.

2.5 Conclusion and Suggestions for Future Research

In the work described in this chapter, we tried to suggest some new approaches for monitoring general profiles combining proper ingredients which have all already been tested and whose good properties have already been shown in literature. These ingredients are, as seen, self-starting charts for getting rid of the estimation of the unknown parameters, the MEWMA/CUSCORE schemes to deal with the accumulation of multivariate observations over time and some nonparametric technique (wavelet transforms or kernel linear smoothing for example) to construct a proper statistic test which efficiently synthesizes the information in the available profiles.

What can be done to improve our study and what could be better investigated in order to improve this research? First of all one could try to better investigate the role of the wavelet transform in the schemes which use them to see if some transforms are appreciably better than other ones and than other methods which do not use wavelets, by trying also other types of wavelet transforms instead of only the two ones used (La10 and Haar wavelets), even if we think that probably the definition of the wavelet transform is not a crucial point. Also the role of the thresholding parameter could be better investigated, in order to try to find a way to search for a criterion to choose the value of k which provides the best performance of a control scheme using wavelets. In the same way one could study the behaviour of the number of equivalent degrees of freedom in the nonparametric chart to try to understand if there is an optimal choice also for it, as we partially did. In this sense, the best thing to do, but it appears very difficult, would be to find a way to make a more "honest" comparison between control charts which use wavelet techniques, SSWFMEWMA and SSWFCUSCORE, and the control chart which uses kernel linear smoothing, SSNEWMA; thus, we should find a way to equal the "quantity of smoothing" provided by the number of equivalent degrees of freedom in kernel linear smoothing to the quantity of smoothing provided by the threshold k in wavelet techniques with Fan's thresholding.

Moreover, this work is limited to profiles data with only one explanatory variable, but it would be interesting to investigate the potentiality of the presented charting schemes also in presence of more than one explanatory variable. Furthermore, we could try to consider to complicate in other ways the reference model for the data, by considering also heteroschedasticity or some kind of dependence in the covariance matrix of the dependence variable or also other probability distributions for the error term, even if it is not so simple to relax the actual assumptions with sequential data.

Chapter 3

A Statistical Test to Assess the Stability of a Profile in Phase I

3.1 A Brief Review on Methods to Assess the Stability of a Process in Phase I

Unfortunately not always in the SPC framework it is possible or convenient to drop the distinction between Phase I and Phase II and, therefore, to use self-starting control charts which use all the data immediately to update the parameter estimates and simultaneously check for OC conditions. In particular self-starting control charts are not very accurate when a shift occurs in the beginning of the process, since in this case we have few IC observations; furthermore they are not so suitable when the shift does not occur from a certain point on, that is when the IC and the OC models cannot be represented together by a change point model.

In these cases, where we have to maintain the distinction between the two phases of the analysis, it is crucial in Phase I to check for IC conditions accurately in order to establish correct control limits for a good Phase II analysis; errors in Phase I may be fatal, since they make us completely miss the IC definition of the process we are studying and compromise the successful performance of the whole SPC analysis.

Even in this specific framework of SPC some charts which monitor processes accurately in Phase I do exist in case of univariate processes, when the underlying process distribution follows some parametric model, more frequently the normal one; there exist some charts in case of profiles too, when the type of profile relation among the variables which rule the process is a known relation, linear or even some specific nonlinear relations, and the distribution of the error follows some parametric distribution, even here most frequently the normal distribution. You can find detailed reviews on Phase I analysis, in particular on univariate Phase I control charts, in Montgomery (2005) and in Chakraborti et al. (2008).

However, it is known that processes and errors are not normal in many applications and very often the statistical properties of commonly employed Phase I control charts, such as the Shewhart-type (Shewhart, 1939), the CUSUM-type (Page, 1955) or the charts based on binary and multiple segmentation (Sullivan & Woodall, 1996; Sullivan, 2002), are highly affected, even for slight deviations, from the specified parametric model.

Since in processes which are not profiles the performances of these parametric Phase I control charts are good only if some parametric assumptions are not violated, one could consider the nonparametric versions of Shewhart, CUSUM or EWMA control charts used for univariate processes and adapt them to a multivariate framework considering profiles. Indeed, in the univariate case, something has been done for Phase I in this direction (for example Jones-Farmer et al. (2009)), but even if these charts behave very well when the process is IC, that is the prescribed IC false alarm probability is conserved regardless of the underlying process distribution, their performance when the process is OC depends on the particular shape of the process distribution and the type of shift from the IC situation. Thus, the choice to adapt them to processes which are profiles seems not to be worthy.

Therefore we would like to propose a new method, completely nonparametric, able to assess the stability of a general and unknown relation between a response variable $\boldsymbol{y}$ and one or more explanatory and deterministic variables $\boldsymbol{x}$ and to recognize OC signals in almost every situation as soon as possible. This new method we are going to propose is actually a kind of multivariate version of an already existing method (Capizzi & Masarotto, 2012) with some slight differences; indeed, their method was applied to a univariate situation in which a sample of m subgroups, each of size n, is collected from the distribution of a quality characteristic $\boldsymbol{y}$.

In addiction, our method will provide a very useful and interesting statistical tool to make some diagnostics in case of instability of the process, that is to try to point out the most probable time interval where the process has gone out of control.

This chapter is arranged in the following way: in Section 3.2 we will expose the reference model and the statistical test we propose in this work; in Section 3.3 we will describe and summarize the results our simulation studies to test performances of this method in different situations; in Section 3.4 we will explain how we can make some diagnostics with this method; finally, in Section 3.5 we will try to give some hints for future research about this topic to improve results obtained in this second part of the thesis.

3.2 Framework and Model

In this framework we assume that a sample of T profiles, each of size n, is collected from the distribution of a quality characteristic, either continuous or discrete, $\boldsymbol{y}$. Let $y_{i,t}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$, be the i^{th} observation of the t^{th} profile such that $\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{n,t})'$ is an n-dimensional vector representing the variable of interest of the considered profile. Differently from Capizzi & Masarotto (2012), we consider also the possibility of a deterministic explanatory variable $\boldsymbol{x}$ since our research is focused on profiles.

When the process is stable, we assume these profiles to be independent and drawn from an unknown but common cumulative distribution function $F_0(\boldsymbol{y}|\boldsymbol{x})$, whereas when the process is unstable[1] the profiles can be thought

[1]With *stable* and *unstable*, in this Phase I SPC framework, we mean the same as *in*

drawn by the following multiple change-point nonparametric model:

$$\boldsymbol{y}_t \sim \begin{cases} F_0(\boldsymbol{y}|\boldsymbol{x}) & \text{if } 0 < t \leq \tau_1 \\ F_1(\boldsymbol{y}|\boldsymbol{x}) & \text{if } \tau_1 < t \leq \tau_2 \\ \dots \\ F_k(\boldsymbol{y}|\boldsymbol{x}) & \text{if } \tau_k < t \leq T\,, \end{cases} \tag{3.1}$$

where $0 < \tau_1 < \tau_2 < \ldots < \tau_k < T$ represent k unknown change points and $F_r(\boldsymbol{y}|\boldsymbol{x})$, $r = 0, \ldots, k$, are unknown cumulative distribution functions for $\boldsymbol{y}$ depending on the values assumed by the explanatory variable $\boldsymbol{x}$.

Notice that model 3.1 includes a great variety of processes which are not stable everywhere. It can perfectly describe processes which present *step* (ex.: model 3.2), *transient* (ex.: model 3.3) and even *isolated* (ex.: model 3.4) shifts.

$$\boldsymbol{y}_t \sim \begin{cases} F_0(\boldsymbol{y}|\boldsymbol{x}) & \text{if } 0 < t \leq \tau_1 \\ F_1(\boldsymbol{y}|\boldsymbol{x}) & \text{if } \tau_1 < t \leq \tau_2 \\ F_2(\boldsymbol{y}|\boldsymbol{x}) & \text{if } \tau_2 < t \leq T \end{cases} \tag{3.2}$$

$$\boldsymbol{y}_t \sim \begin{cases} F_0(\boldsymbol{y}|\boldsymbol{x}) & \text{if } 0 < t \leq \tau_1 \\ F_1(\boldsymbol{y}|\boldsymbol{x}) & \text{if } \tau_1 < t \leq \tau_2 \\ F_0(\boldsymbol{y}|\boldsymbol{x}) & \text{if } \tau_2 < t \leq T \end{cases} \tag{3.3}$$

$$\boldsymbol{y}_t \sim \begin{cases} F_0(\boldsymbol{y}|\boldsymbol{x}) & \text{if } t \neq \tau \\ F_1(\boldsymbol{y}|\boldsymbol{x}) & \text{if } t = \tau \end{cases} \tag{3.4}$$

Now, let us explain more in detail how we constructed this multivariate test, on the basis of the previous univariate version presented in Capizzi & Masarotto (2012). Since we are interested to suggest a new type of Phase I analysis, we would like to provide a final statistical test in order to obtain a p-value which will test the hypotheses

$$\begin{cases} H_0\colon & \text{the process is stable } \forall t,\ t = 1, \ldots, T \\ H_1\colon & \exists\, t,\ t = 1, \ldots, T\colon \text{the process is not stable}\,. \end{cases}$$

control and *out of control* respectively.

Furthermore we would like to provide a diagnostic tool, connected with the statistical test, able to identify in some way time and type of changes when we reject the hypothesis of stability (see Section 3.4).

The reference model for our study will be

$$\boldsymbol{y}_t = \mu_0(\boldsymbol{x}) + \delta(\boldsymbol{x})\mathbb{I}_{[\tau_1,\tau_2]}(t) + \boldsymbol{\epsilon}_t\,, \quad t = 1, \ldots, T\,, \tag{3.5}$$

where you can see that the OC deviation $\delta(\boldsymbol{x})$ occurs from the τ_1^{th} profile to the τ_2^{th} one, which thus lasts for an interval of $\ell = \tau_2 - (\tau_1 - 1) = \tau_2 - \tau_1 + 1$ time units. The number of observations T is fixed; in simulation studies (Section 3.3) you will find in detail which distributions and which covariance structures we will consider for the error term $\boldsymbol{\epsilon}_t$.

Note that in this second part of this thesis, assumptions are more relaxed with respect to the first part (see Section 2.2 at page 11): we do not restrict to the case of Gaussian, homoschedastic and independent errors as done before and this can be done because when profiles are not sequential, that is when time T is fixed and not undefined, it is more simple to treat them and thus to consider also more complicated models. The only assumption the method will make on the data $\boldsymbol{y}$ is the independence between the profiles, because we would like to work in an almost completely nonparametric way and to construct a statistical tool valid in almost every possible situation.

This final statistical test is based on the cumulated sums of the data

$$\boldsymbol{S}_{\bar{t}} = \sum_{t=1}^{\bar{t}} \boldsymbol{y}_t\,, \quad \bar{t} = 1, \ldots, T\,.$$

In particular for $\bar{t} = T$, we get the vector of the totals

$$\boldsymbol{S}_T = \sum_{i=1}^{T} \boldsymbol{y}_i = T\overline{\boldsymbol{y}}\,,$$

which is, indeed, the last cumulated sum.

Let us suppose that we are interested to evaluate the stability of the profiles in particular between the time intervals from $t = \tau_1$ to $t = \tau_2$ and the rest of the whole time interval, that is from $t = 1$ to $t = \tau_1 - 1$ jointly with

$t = \tau_2 + 1$ to $t = T$; we are thus interested if the distribution which generates the data is always the same in the two time intervals considered. A possible measure of the stability of the profiles between the two time intervals can be the vector of the difference of the sample means between the two time intervals, $\overline{\boldsymbol{y}}_{[\tau_1,\tau_2]} - \overline{\boldsymbol{y}}_{\overline{[\tau_1,\tau_2]}}$, which can be written (see the appendix at page 69) in the following way:

$$\overline{\boldsymbol{y}}_{[\tau_1,\tau_2]} - \overline{\boldsymbol{y}}_{\overline{[\tau_1,\tau_2]}} = \boldsymbol{T}_{\tau_1,\tau_2} = \frac{\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)} - (\tau_2 - \tau_1 + 1)\overline{\boldsymbol{y}}}{(\tau_2 - \tau_1 + 1)\left(1 - \frac{\tau_2-\tau_1+1}{T}\right)} . \tag{3.6}$$

The higher the magnitude of $\boldsymbol{T}_{\tau_1,\tau_2}$ is, the more unstable the two (actually three, but the first and the third are thought connected) time intervals are.

Since we want to work in a completely nonparametric point of view, in order to try to be able to recognize every kind of instability in the profiles, we smooth the statistic $\boldsymbol{T}_{\tau_1,\tau_2}$ with a kernel linear smoothing:

$$\boldsymbol{K}_{\tau_1,\tau_2,h} = \boldsymbol{S}_h \boldsymbol{T}_{\tau_1,\tau_2} ,$$

where $\boldsymbol{S}_h$ is the kernel smoothing matrix, defined as for SSNEWMA in Section 2.2, and h is a proper bandwidth of the smoother (we can also choose the degrees of freedom and not the bandwidth). The next step is to calculate a standardized norm of the vector $\boldsymbol{K}_{\tau_1,\tau_2,h}$ through the quadratic form

$$T_{\tau_1,\tau_2,h} = \boldsymbol{K}'_{\tau_1,\tau_2,h} \boldsymbol{V}_h \boldsymbol{K}_{\tau_1,\tau_2,h} ,$$

where $\boldsymbol{V}_h = \boldsymbol{S}_h + \boldsymbol{S}'_h - \boldsymbol{S}'_h \boldsymbol{S}_h$, up to the variance of the sample difference $\overline{\boldsymbol{y}}_{[\tau_1,\tau_2]} - \overline{\boldsymbol{y}}_{\overline{[\tau_1,\tau_2]}}$, is the inverse of the covariance matrix of $\boldsymbol{K}_{\tau_1,\tau_2,h}$. We will completely standardize our statistic in the next steps.

The IC probability distribution function of all $T_{\tau_1,\tau_2,h}$, $\tau_1 = 1, \ldots, T$, $\tau_2 = 2, \ldots, T$, $\tau_1 < \tau_2$, $h = 1, \ldots, H$, depends on the unknown distribution $F_0(\boldsymbol{y}|\boldsymbol{x})$ and thus is unknown as well, but with some tricks we can obtain a p-value anyway: let $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$ be the $n \times T$ matrix of the profiles and let $\mathcal{S}$ be the set of all the $T!$ $n \times T$ matrices obtainable by permuting the columns of $\boldsymbol{Y}$. Then, it is known (for further information on permutation methods, see Pesarin (2001)) that under H_0

$$\Pr\left[\boldsymbol{Y} = a|\mathcal{S}\right] = \begin{cases} \dfrac{1}{T!} & \text{if } a \in \mathcal{S} \\ 0 & \text{if } a \notin \mathcal{S}\,. \end{cases} \tag{3.7}$$

Since the previous result does not depend on the IC distribution $F_0(\boldsymbol{y}|\boldsymbol{x})$, given a test statistic, we can compute a p-value, which is conditioned only on the set of permutations of the profiles $\mathcal{S}$. This p-value is calculated as the proportion of permutations where the value of the statistic is greater or equal to the value of the statistic computed on the original sample. We are interested only in permuting times and not the whole pooled sample, as in Capizzi & Masarotto (2012), which presents this permutation method referring to the global order statistic of the pooled sample of all nT observations.

In particular, knowing that the number of permutations $T!$ in general is huge and that, in our case, for each of the H chosen cases of number of degrees of freedom we want to consider, we have $\frac{T(T-1)}{2}$ test statistics $T_{\tau_1,\tau_2,h}$ (it is the number of combinations of the values τ_1 and τ_2), each assuming that between τ_1 and τ_2 the process is unstable and outside is stable (therefore we have $H\frac{T(T-1)}{2}$ test statistics all together), we suggest to generate L random permutations of the data. Then for each permutation l, $l = 1, \ldots, L$, we can calculate the value of $T_{\tau_1,\tau_2,h}$, namely $\widetilde{T}^{(l)}_{\tau_1,\tau_2,h}$. At this point we standardize both the statistics $T_{\tau_1,\tau_2,h}$, computed on the original sample, and the group of statistics $\widetilde{T}^{(l)}_{\tau_1,\tau_2,h}$, each computed on the l^{th} permutation, in the following way:

$$Z_{\tau_1,\tau_2,h} = \frac{T_{\tau_1,\tau_2,h} - u_{\tau_1,\tau_2,h}}{v_{\tau_1,\tau_2,h}} \quad \text{and} \quad \widetilde{Z}^{(l)}_{\tau_1,\tau_2,h} = \frac{\widetilde{T}^{(l)}_{\tau_1,\tau_2,h} - u_{\tau_1,\tau_2,h}}{v_{\tau_1,\tau_2,h}}\,, \tag{3.8}$$

where

$$u_{\tau_1,\tau_2,h} = \frac{1}{L}\sum_{l=1}^{L}\widetilde{T}^{(l)}_{\tau_1,\tau_2,h} \quad \text{and} \quad v^2_{\tau_1,\tau_2,h} = \frac{1}{L-1}\sum_{l=1}^{L}\left(\widetilde{T}^{(l)}_{\tau_1,\tau_2,h} - u_{\tau_1,\tau_2,h}\right)^2$$

are the sample mean and variance of the statistics $\widetilde{T}^{(l)}_{\tau_1,\tau_2,h}$ respectively.

Finally we compute the following overall control statistics for both the original sample and the permutations:

$$W = \max_{\tau_1,\tau_2,h} Z_{\tau_1,\tau_2,h} \text{ and } \widetilde{W}^{(l)} = \max_{\tau_1,\tau_2,h} \widetilde{Z}^{(l)}_{\tau_1,\tau_2,h} . \tag{3.9}$$

Note that the standardization of $T_{\tau_1,\tau_2,h}$ and $\widetilde{T}^{(l)}_{\tau_1,\tau_2,h}$ with $u_{\tau_1,\tau_2,h}$ and $v_{\tau_1,\tau_2,h}$ is crucial, since we need to have comparable quantities to obtain an honest maximum among them.

Permutations are thus useful to simulate lots of IC values of the statistic W and therefore a natural way to provide a conditioned p-value can be defined as

$$p = \frac{1}{L}\sum_{l=1}^{L} \mathbb{I}\Big(\widetilde{W}^{(l)} \geq W\Big) . \tag{3.10}$$

As the natural definition of a p-value suggests, it is defined as the proportion of cases where the values of the statistic W computed on the permutations of the data (representing a sample of theoretical values of W when the process is stable everywhere) exceeds or is equal than the value of W observed on the original data.

3.3 Simulation Study

In this section we investigate the performance of the test in different situations; for simulations we are going to describe we used a code written in the `R` language (R Development Core Team, 2010).

Before describing simulation studies in detail, we have to make some clarifications: differently from what we said in the previous section, in these simulation studies we apply some tricks. First the number of test statistics among which we choose the maximum, is not $H\frac{T(T-1)}{2}$, but $H\Big[6(T-9)+\frac{(T-10)(T-9)}{2}\Big]$ because we will actually consider not all combinations of possible τ_1 and τ_2, but only those where both ℓ and $T-\ell$ are at least 5, that is both the IC and the OC periods have at least 5 observations. Furthermore we decided to consider, for simplicity, only $H = 3$ choices of bandwidths/degrees of freedom, with $g_1 = 3.398$ ($df = 3$), $g_2 = 1$ ($df = 6.59$) and $g_3 = 0.682$ ($df = 9$) and anyway we will see that the choice of g (or df) is not crucial. In simulation studies, presented in detail in the following lines, we will apply these two last tricks.

Let us specify the reference model 3.5 at page 49 introduced in the previous section more particularly for our simulation studies:

- about the error term: $\boldsymbol{\epsilon}_t = \boldsymbol{A}_k \boldsymbol{\epsilon}_t^*$, where $\boldsymbol{A}_k$, $k = 1, 2, 3$, is the Cholesky factor of a specific covariance matrix $\boldsymbol{\Sigma}_k$ and for ϵ_{tj}^*, $j = 1, \ldots, n$, the components of $\boldsymbol{\epsilon}_t^*$, we consider three distributions: $\mathrm{N}(0,1)$, t_3 or $\mathrm{SN}(0,1,3)$;

- about the OC shift: $\delta(\boldsymbol{x}) = \alpha[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]$; the shift is rearranged such that by varying the intra-profile correlation, it has always the same meaning;

- $\alpha = \sqrt{\dfrac{[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]'[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]}{[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]'\boldsymbol{\Sigma}_k{}^{-1}[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]}}$ is needed to make different scenarios comparable and it has been calculated such that $\alpha^2[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]'\boldsymbol{\Sigma}_k{}^{-1}[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})] = [\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]'[\mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})]$;

- $\ell = \tau_2 - (\tau_1 - 1) = \tau_2 - \tau_1 + 1$ is the length of the instability period of the process;

- the number of observations T is fixed;

- $\mathrm{Var}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}_k = \sigma_r^2[\sigma_{ij,k}]$, where $\sigma_{ij,k} = \begin{cases} \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases} & \text{if } k = 1 \\ \rho^{|i-j|} & \text{if } k = 2 \\ \begin{cases} 1 \text{ if } i = j \\ \lambda \text{ if } i \neq j \end{cases} & \text{if } k = 3\,; \end{cases}$ is the (i, j) element of the covariance matrix $\boldsymbol{\Sigma}_k$; σ_r^2, $r = 1, 2, 3$, is the variance of the distribution of each component of $\boldsymbol{\epsilon}_t^*$: 1 with $\mathrm{N}(0,1)$ ($r = 1$), 3 with t_3 ($r = 2$) and 0.427 with $\mathrm{SN}(0,1,3)$ ($r = 3$). Thus, we consider three possibilities of intra-profile correlation: uncorrelation ($k = 1$); cascade ($k = 2$), that is the higher the distance of the components, the more the correlation decreases; correlation always equal to a same value λ ($k = 3$). We will take $\rho = 0.6$ and $\lambda = 0.2$.

First of all we checked that this test works well when the process is IC: we considered the three different distributions of the error, the three covariance matrices and two values of the number of time observations T (10 and 20): we obtain $3 \times 3 \times 2 = 18$ IC scenarios all together. For each IC scenario we calculate 100 p-values of the test, each computed using $L = 1000$ permutations. The number of p-values for each scenario is not huge, but the code which executes the simulations is really slow and even in this way each scenario takes days to be simulated completely; furthermore, in any case, already with 100 p-values it is possible to have an idea of their distribution and to recognize, if the method is good, if the analyzed scenario is effectively IC or OC (see figures 3.3 and 3.5).

After realizing that IC the method works well we tested this method in OC situations in two ways.

First we tried to explore performances of our test in 162 different scenarios, obtained by the 27 combinations of before (actually the 18 combinations of before plus other 9 combinations provided when $T = 50$) applied to 6 situations ($27 \times 6 = 162$) already used in Chapter 2: OC1A, OC1E, OC2A, OC2E, OC3A and OC3E (for each of the three OC models we chose the scenario with the smallest value of the OC parameter ν and the scenario with the medium value of ν: see table 2.1 at page 20 for details on these OC situations). In all scenarios, the IC/OC pattern is always the same: the first $\frac{T}{2}$ observations are IC, whereas from the $\left(\frac{T}{2}+1\right)^{\text{th}}$ to the T^{th} are OC. We provided only one p-value for each of the 162 scenarios just to explore the capabilities of this test in different situations.

Then we chose one situation, OC3E, to be investigated a bit more deeply: we considered the three distributions of the error of before, only the first covariance matrix of before, only 20 as number of time observations T (10 is not so interesting and more than 20 time observations take too much long, as you will see in Section 3.5) and 4 different IC/OC patterns. We thus obtain $3 \times 1 \times 1 \times 4 = 12$ specific scenarios in this second group of OC simulations, which will be used also to make some diagnostics (see Section 3.4). As in IC simulations, even here we simulate 100 p-values for each scenario and each p-value is calculated using $L = 1000$ permutations. The 4 patterns of before are generated in the following way:

1. 1^{th} pattern: 10 IC observations and 10 OC observations;
2. 2^{nd} pattern: 15 IC observations and 5 OC observations;
3. 3^{rd} pattern: 5 IC observations, 10 OC observations and finally 5 IC observations again;
4. 4^{th} pattern: 10 IC observations, 5 OC observations and finally 5 IC observations again.

We explored more deeply one OC situation, in this case model OC3E, because we intended to explore performances of our test in different situations: both in step OC shifts (patterns 1 and 2) and in transient OC shifts (patterns 3 and 4); and also both when the OC period is as long as the IC period (patterns 1 and 3) and when the IC period is longer than the OC one (pattern 2 and 4).

Results are quite encouraging in all three simulation studies: let us see them a bit more in detail.

In IC scenarios, we expect p-values to be uniformly distributed. In fig. 3.1 you can find an example of distribution of the 100 simulated p-values in an IC situation; here, in particular, we considered the case of normal errors with the second type of covariance matrix and $T = 20$. Our method, in this IC situation, seems to work quite well since the simulated distribution of p-values seems to be uniform. The boxplot (fig. 3.2) and the descriptive statistics (table 3.1) about the simulated p-values in this IC situation confirm our impression got from the histogram.

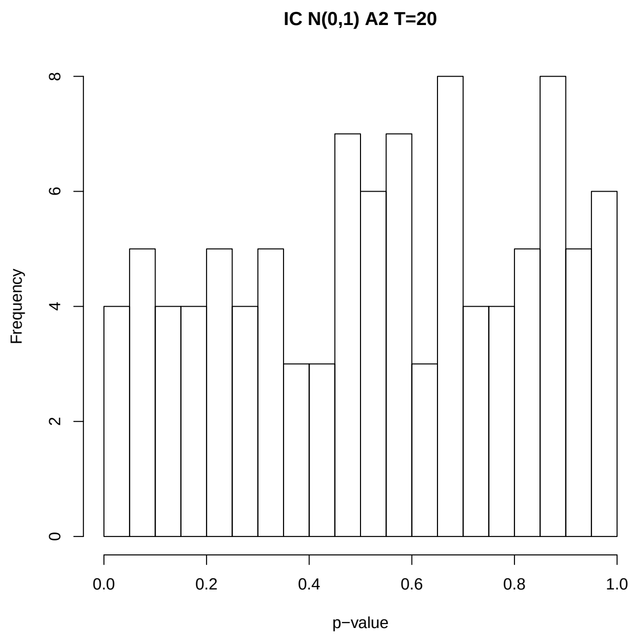

Figure 3.1: Simulated distribution of 100 p-values in the considered IC situation.

When the significance level α is declared equal to 0.05, the estimated first type error is 0.04; when the nominal α is 0.01, the estimated one is 0.01 too and when the nominal α is 0.1, the estimated one is 0.09. These results are rather good and are a further confirm that our test seems to behave very well in this IC situation.

Min	1st Q	Median	Mean	3st Q	Max	Skewness	SD
0.0090	0.2952	0.5505	0.5341	0.7732	0.9980	-0.1758	0.2876

Table 3.1: Descriptive statistics of p-values in the considered IC situation.

Now let us briefly consider the first group of OC simulations: it is just explorative, indeed we provided lots of OC situations (162), but only one p-value for each situation, since the code is too slow to provide a huge number of p-values for a great number of scenarios. Since it is not possible to report here all 162 p-values of example, we limit to say that results are quite

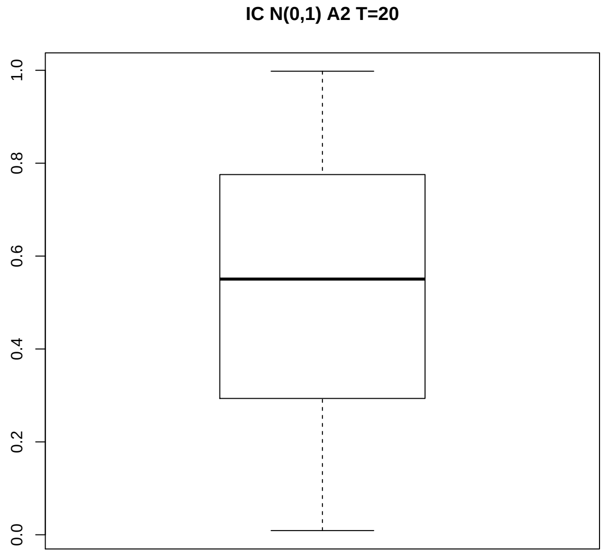

Figure 3.2: Boxplot of 100 p-values in the considered IC situation.

encouraging, but not so good: only one p-value out of 3, more or less, is significant, but we cannot forget that we are evaluating very small deviations distributed on whole profiles. It is thus probably difficult to do better than this. No particular ingredient of these scenarios (covariance matrix and distribution of the errors, time observations, type of OC deviation from the IC model) seem to be particularly favoured with respect to the other ones.

Finally let us consider the second, more interesting, group of OC simulations: in fig. 3.3 you can find an example of distribution of the 100 simulated p-values in an OC situation with respect to the second group of OC simulations; here, in particular, we considered the fourth pattern applied to the case of errors with a skew normal distribution; in the other 11 considered scenarios results are quite similar. As you can see, in this case the method works quite well since the distribution of p-values is really skew towards 0, which means that in general our test is able to signal that some profiles are not IC. The boxplot gives a similar vision of the distribution of the p-values:

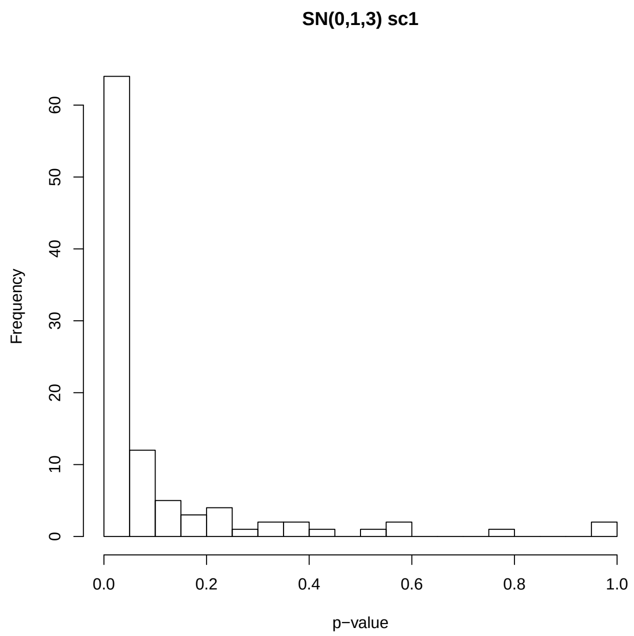

Figure 3.3: Simulated distribution of 100 p-values in the considered OC situation.

see fig. 3.4.

Some descriptive statistics about these p-values are reported in table 3.2. The estimated power $1-\beta$ when the significance level α is equal to 0.05 is 0.56; furthermore when $\alpha = 0.01$, $1 - \beta = 0.35$ and when $\alpha = 0.1$, $1 - \beta = 0.67$.

Min	1st Q	Median	Mean	3st Q	Max	Skewness	SD
0.0000	0.0058	0.0315	0.1552	0.1672	1.0000	2.2035	0.2644

Table 3.2: Descriptive statistics of p-values in the considered OC situation.

Moreover, in order to explore the effect of the choice of the bandwidth, we report the distributions of the p-values in function of the chosen values for the parameter g of the bandwidth ($g_1 = 3.398$ ($df = 3$), $g_2 = 1$ ($df = 6.59$) and $g_3 = 0.682$ ($df = 9$)); these p-values are obtained from the "partial" control statistics $W_h = \max_{\tau_1,\tau_2} Z_{\tau_1,\tau_2,h}$. We compare these three distributions with the distribution of p-values of before, that is the one obtained from the final

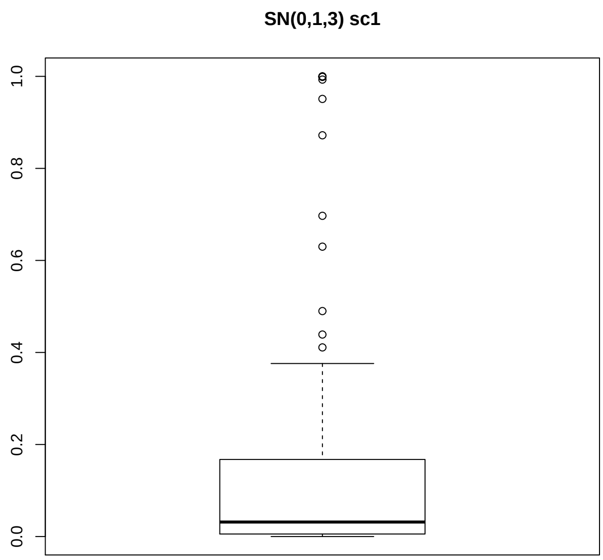

Figure 3.4: Boxplot of 100 p-values in the considered OC situation.

statistic $W = \max_{\tau_1,\tau_2,h} Z_{\tau_1,\tau_2,h}$, which maximizes also with respect to the three considered values of h.

Look at fig. 3.5: as you can see the choice of the bandwidth seems not to be so crucial and maybe it would be even not necessary to maximize the statistics with respect to some possible bandwidths, but simply to choose one of them and use it. In any case, by doing the maximization we protect ourselves from completely wrong choices of the bandwidth. In table 3.3 we sum up some descriptive statistics of the distributions of p-values in function of g.

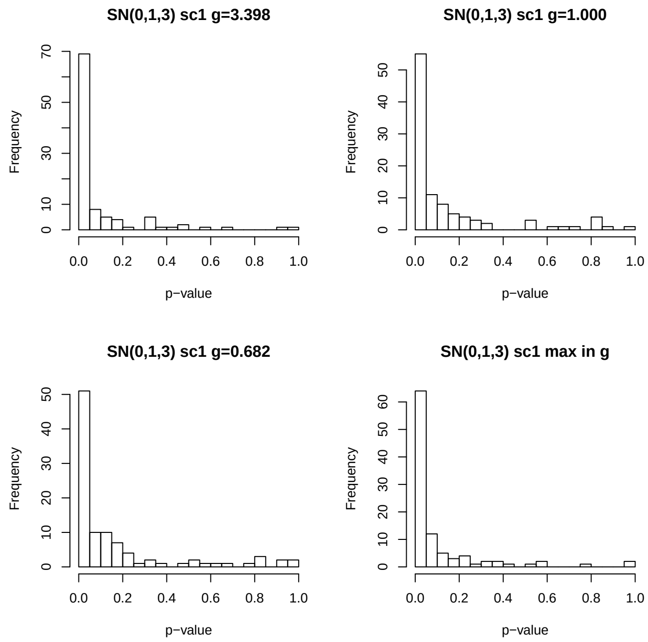

Figure 3.5: Simulated distributions of 100 p-values in the considered OC situation with respect to different bandwidths.

g	Min	1st Q	Median	Mean	3st Q	Max	Skewness	SD
3.398	0.0000	0.0058	0.0285	0.1494	0.1952	0.9980	2.2912	0.2540
1.000	0.0000	0.0118	0.0425	0.2011	0.2585	1.0000	1.6056	0.2930
0.682	0.0000	0.0080	0.0500	0.2142	0.3340	1.0000	1.4338	0.2935
max in g	0.0000	0.0058	0.0315	0.1552	0.1672	1.0000	2.2035	0.2644

Table 3.3: Descriptive statistics of p-values in the considered OC situation.

3.4 Only Assessing the Stability or even some Diagnostics?

As we have previously said, we would like to provide a diagnostic tool together with this statistical test for trying to recognize which are the most probable values for τ_1 and τ_2, that is where the process could be OC when we refuse the null hypothesis of stability of the process. We simply suggest to consider the most probable period of instability the time interval $[\tau_1, \tau_2]$ where we obtain the statistic W, that is where $Z_{\tau_1,\tau_2,h}$ is maximum.

To see if this method works, let us take the second group of OC simulations and consider, for each scenario, the $M = 100$ simulated values of W and represent in a special graph the correspondent M points (τ_1, τ_2) where they maximize the statistics $Z_{\tau_1,\tau_2,h}$. In the next pages you find some examples of this graphs, where the darker each point (τ_1, τ_2) is, the higher the frequency of that point. We chose to report here in particular the following scenarios:

- all 4 OC scenarios (see the previous section for the definitions of the scenarios) with errors distributed like a Student's t (figg. 3.6-3.9);
- the third OC scenario with normal errors (fig. 3.10);
- the third OC scenario with skew normal errors (fig. 3.11).

Furthermore we chose an example of IC situation too, in particular the one with normal errors, second type of covariance matrix and $T = 20$ (fig. 3.12), to see how this diagnostic tool of our test behaves when the process is always stable.

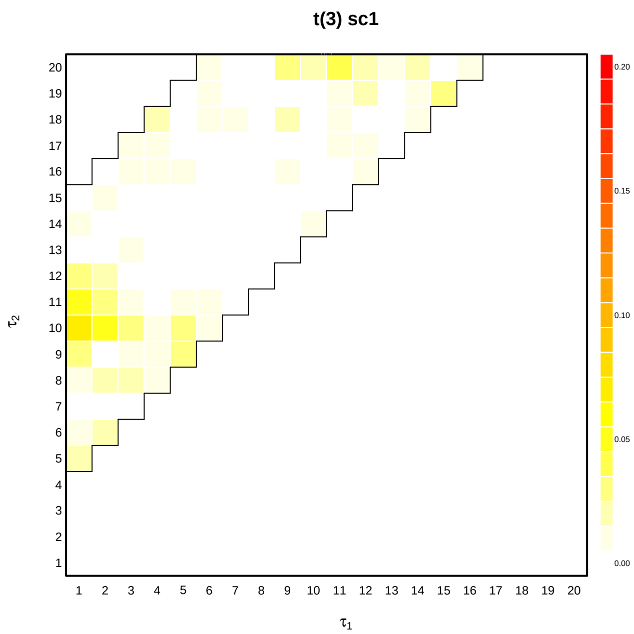

Figure 3.6: Distribution of points (combination of times) where the statistic has the maximum value in the considered OC situation.

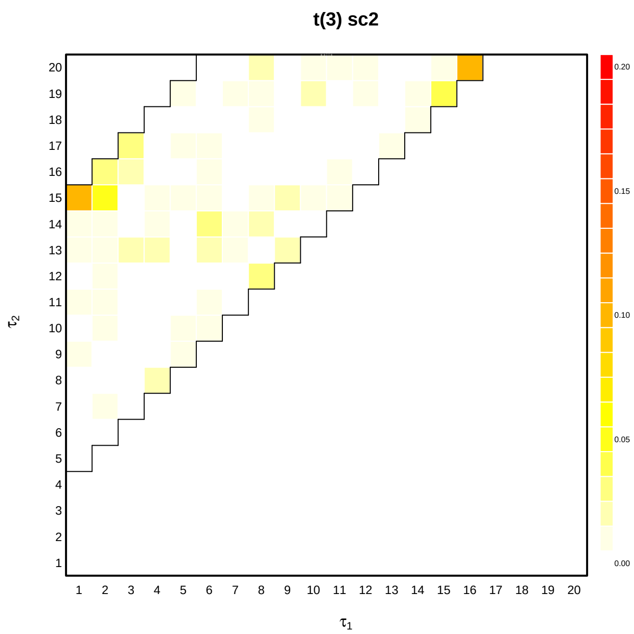

Figure 3.7: Distribution of points (combination of times) where the statistic has the maximum value in the considered OC situation.

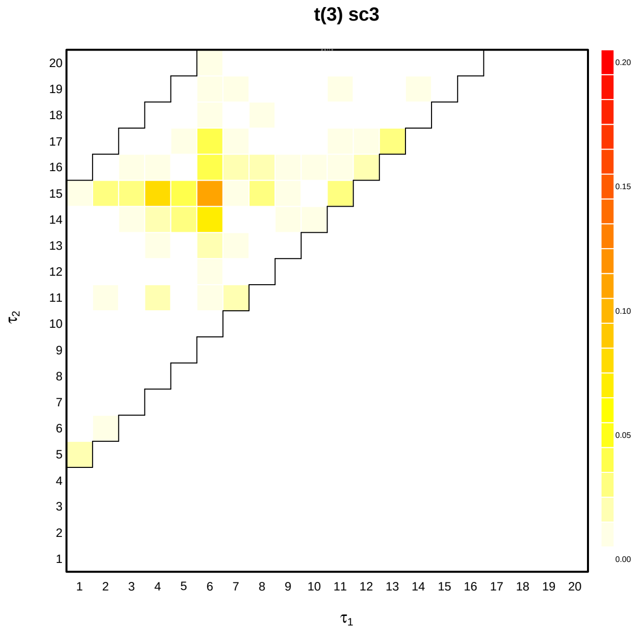

Figure 3.8: Distribution of points (combination of times) where the statistic has the maximum value in the considered OC situation.

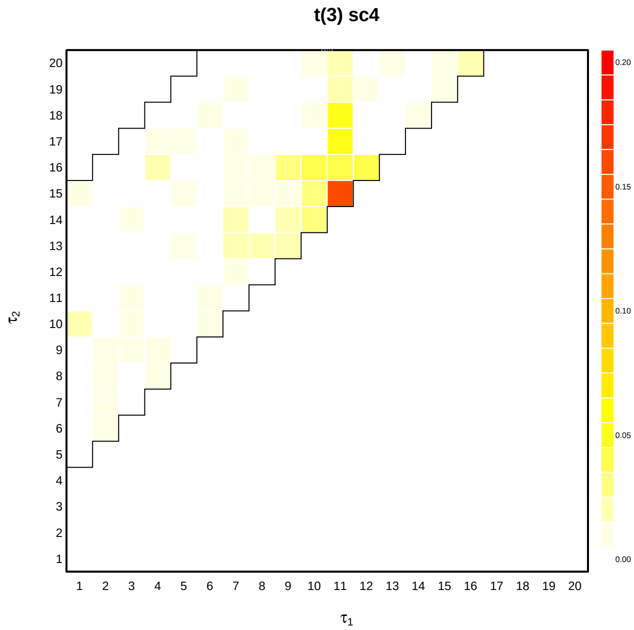

Figure 3.9: Distribution of points (combination of times) where the statistic has the maximum value in the considered OC situation.

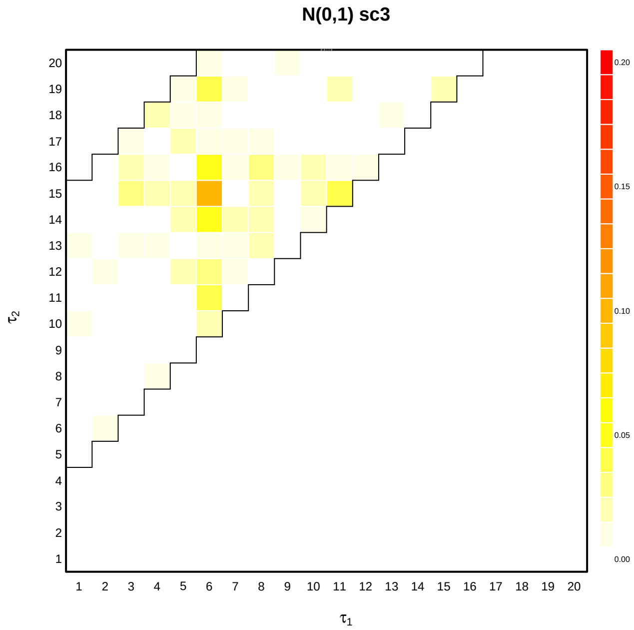

Figure 3.10: Distribution of points (combination of times) where the statistic has the maximum value in the considered OC situation.

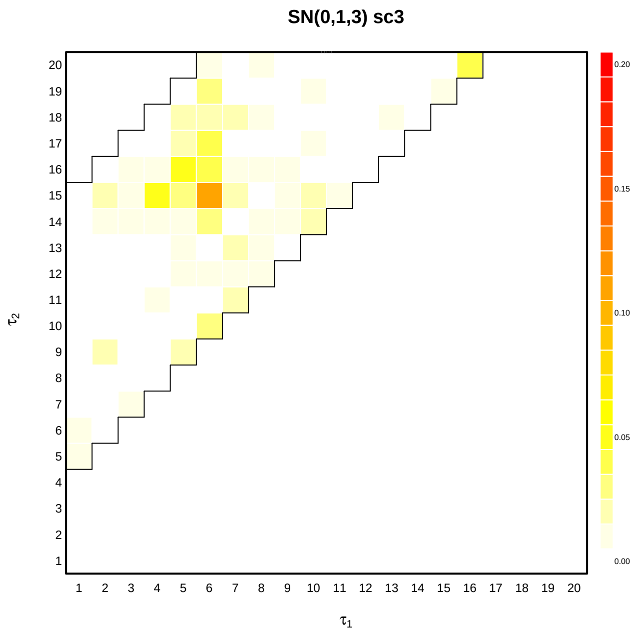

Figure 3.11: Distribution of points (combination of times) where the statistic has the maximum value in the considered IC situation.

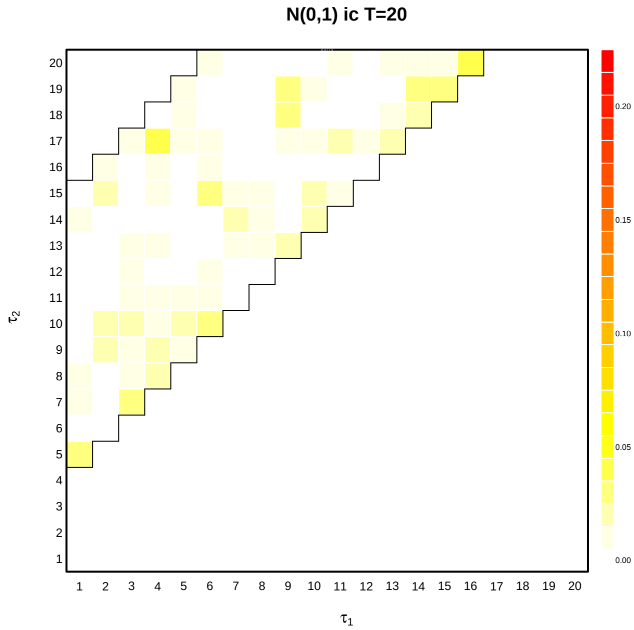

Figure 3.12: Distribution of points (combination of times) where the statistic has the maximum value in the considered IC situation.

Let us consider the four scenarios with errors drawn from a Student's t distribution (figg. 3.6-3.9). You can see that the instability in the first scenario, where $\tau_1 = 11$ and $\tau_2 = 20$ (but the method could exchange the IC and the OC period and think they are exactly the contrary and so it could also tell us that $\tau_1 = 1$ and $\tau_2 = 10$), is more difficult to grasp: in fact here the IC and the OC periods are compact (step deviation) and have the same length. The instability periods of the second and the third scenarios are grasped quite well: this happens probably because the OC period is longer than the IC one (scenario 2) or because the OC period is in the middle of the process, which is thus transient (scenario 3). The instability of the fourth scenario seems to be grasped very well, probably since the OC period is very short and in the middle of the process.

We report scenario 3 also for normal (fig. 3.10) and skew normal (fig. 3.11) errors: their distribution seems not to be a crucial point; the only difference is that with normal errors, deviations from the location of the exact point (τ_1, τ_2) seem to spread in every direction, differently from errors with other distributions. The same happens also for the other scenarios.

Finally we can see also that in the IC example (fig. 3.12) our diagnostic tool seems to behave well, since there are no "accumulation points" and, therefore, no information on a more probable (τ_1, τ_2) seem to be signaled out.

Note that all graphs have a domain ladder shaped and this because, as we have already said before, we chose to consider only cases when both the stable and the unstable periods are longer or equal than 5 time units.

3.5 Conclusion and Suggestions for Future Research

In the work described in this chapter we tried to introduce a new almost completely nonparametric model for Phase I analysis of profile data.

Probably it is possible to do something more in this research field. First we could envelope a statistical test which is useful to treat also profile data with other types of OC time intervals, for example more than one OC interval (step shifts from the IC model) or an OC interval consisting in only one time observation (isolated shift from the IC model).

Furthermore the code which applies this method to profile data is still really slow and probably one could improve it: nowadays a general scenario with $L = 1000$ permutations, $M = 100$ simulations (to obtain "only" 100 p-values) takes almost 2 hours when $T = 10$, 34 hours when $T = 20$ and even almost 12 days and a half when $T = 50$! By improving the code, trying to increase its speed, it will be possible to explore performances of this method more in detail.

Finally, the diagnostic tool seems to work quite well and it could be a good proposal to try to find the most probable instability period $[\tau_1, \tau_2]$.

Appendix

Proof 1: a good form of the statistic $\boldsymbol{T}_{\tau_1,\tau_2}$

$$
\begin{aligned}
\boldsymbol{T}_{\tau_1,\tau_2} &= \overline{\boldsymbol{y}}_{[\tau_1,\tau_2]} - \overline{\boldsymbol{y}}_{\overline{[\tau_1,\tau_2]}} \\
&= \frac{\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)}}{\tau_2 - \tau_1 + 1} - \frac{T\overline{\boldsymbol{y}} - (\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)})}{T - (\tau_2 - \tau_1 + 1)} \\
&= \frac{[T - (\tau_2 - \tau_1 + 1)](\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)}) - (\tau_2 - \tau_1 + 1)[T\overline{\boldsymbol{y}} - (\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)})]}{(\tau_2 - \tau_1 + 1)[T - (\tau_2 - \tau_1 + 1)]} \\
&= \frac{T(\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)}) - (\tau_2 - \tau_1 + 1)(\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)}) - T(\tau_2 - \tau_1 + 1)\overline{\boldsymbol{y}}}{(\tau_2 - \tau_1 + 1)[T - (\tau_2 - \tau_1 + 1)]} + \\
&\quad + \frac{(\tau_2 - \tau_1 + 1)(\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)})}{(\tau_2 - \tau_1 + 1)[T - (\tau_2 - \tau_1 + 1)]} \\
&= \frac{T(\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)}) - T(\tau_2 - \tau_1 + 1)\overline{\boldsymbol{y}}}{(\tau_2 - \tau_1 + 1)[T - (\tau_2 - \tau_1 + 1)]} \\
&= T\frac{\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)} - (\tau_2 - \tau_1 + 1)\overline{\boldsymbol{y}}}{(\tau_2 - \tau_1 + 1)[T - (\tau_2 - \tau_1 + 1)]} \\
&= \frac{\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)} - (\tau_2 - \tau_1 + 1)\overline{\boldsymbol{y}}}{(\tau_2 - \tau_1 + 1)\dfrac{T - (\tau_2 - \tau_1 + 1)}{T}} \\
&= \frac{\boldsymbol{S}_{\tau_2} - \boldsymbol{S}_{(\tau_1-1)} - (\tau_2 - \tau_1 + 1)\overline{\boldsymbol{y}}}{(\tau_2 - \tau_1 + 1)\left(1 - \dfrac{\tau_2 - \tau_1 + 1}{T}\right)}
\end{aligned}
$$

Bibliography

Akhavan Niaki, S., Abbasi, B. & Arkat, J. (2007). A generalized linear statistical model approach to monitor profiles. *International Journal of Engineering* 3 233–242.

Brill, R. V. (2001). A case study for control charting a product quality measure that is a continuous function over time. In *Presentation at the 45th Annual Fall Technical Conference*. Toronto, Ontario.

Capizzi, G. & Masarotto, G. (2003). An adaptive Exponentially Weighted Moving Average control chart. *Technometrics* 45 199–207.

Capizzi, G. & Masarotto, G. (2008). Practical design of generalized likelihood ratio control charts for autocorrelated data. *Technometrics* 50 357–370.

Capizzi, G. & Masarotto, G. (2010a). Linear profile monitoring using an adaptive EWMA control chart. *2010 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* 1425–1429.

Capizzi, G. & Masarotto, G. (2010b). Self-starting CUSCORE control charts for individual multivariate observations. *Journal of Quality Technology* 42 136–151.

Capizzi, G. & Masarotto, G. (2011). A least angle regression control chart for multidimensional data. *Technometrics* 53 285–296.

Capizzi, G. & Masarotto, G. (2012). Phase I distribution-free analysis of univariate data. *(still to be published)* .

Chakraborti, S., Human, S. W. & Graham, M. A. (2008). Phase i statistical process control charts: An overview and some results. *Quality Engineering* 21 52–62.

Chicken, E., Pignatiello, J. & Simpson, J. (2009). Statistical process monitoring of nonlinear profiles using wavelets. *Journal of Quality Technology* 41 198–212.

Colosimo, B. M. & Pacella, M. (2008). On the use of principal component analysis to identify systematic patterns in roundness profiles. *Quality and Reliability Engineering International* 23 707–725.

Doroslovački, M. L. (1998). On the least asymmetric wavelets. *Signal Processing, IEEE Transactions on* 46 1125 –1130.

Fan, J. (1996). Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association* 91 674–688.

Fan, J. & Lin, S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association* 93 1007–1021.

Han, D. & Tsung, F. (2006). A reference-free CUSCORE chart for dynamic mean change detection and unified framework for charting performance comparison. *Journal of American Statistical Association* 101 368–386.

Hawkins, D. M. (1987). Self-starting cusum charts for location and scale. *The Statistician* 36 299–315.

Hawkins, D. M. & Maboudou-Tchao, E. M. (2007). Self-starting multivariate Exponentially Weighted Moving Average control charting. *Technometrics* 49 199–209.

Jensen, W. A. & Birch, J. B. (2009). Profile monitoring via nonlinear mixed models. *Journal of Quality Technology* 41 18–34.

Jeong, M. K. & Lu, J.-C. (2006). Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research* 44 729–744.

Jin, J. & Shi, J. (1999). State space modeling of sheet metal assembly for dimensional control. *ASME Transactions, Journal of Manufacturing Science and Engineering* 121 756–762.

Jones-Farmer, L. A., Jordan, V. & Champ, C. W. (2009). Distribution-free phase i control charts for subgroup location. *Journal of Quality Technology* 41 304–316.

Kang, L. & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology* 32 418–426.

Kim, K., Mahmoud, M. A. & Woodall, W. H. (2003). On the monitoring of linear profiles. *Journal of Quality Technology* 35 317–328.

Mahmoud, M. A., Parker, P. A., Woodall, W. H. & Hawkins, D. M. (2007). A change-point method for linear profile data. *Quality and Reliability Engineering International* 23 247–268.

Mahmoud, M. A. & Zahran, A. R. (2010). A multivariate adaptive exponentially weighted moving average control chart. *Communications in Statistics - Theory and Methods* 39 606–625.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*. New York: Wiley, 5th ed.

Mosesova, S., Chipman, H., MacKay, R. J. & Steiner, S. (2006). Profile monitoring using mixed-effects models. Technical report, University of Waterloo.

NASON, G. P. (2008). *Wavelet methods in Statistics with R*. Springer-Verlag.

NOOROSSANA, R., SAGHAEI, A. & AMIRI, A. (2011). *Statistical Analysis of Profile Monitoring*. New York: Wiley.

PAGE, E. S. (1955). A test for change in a parameter occurring at an unknown point. *Biometrika* 42 523–527.

PESARIN, F. (2001). *Multivariate Permutation Tests: with Applications in Biostatistics*. New York: Wiley.

QIU, P. & ZOU, C. (2010). Control chart for monitoring nonparametric profiles with arbitrary design. *Statistica Sinica* 20 1655–1682.

R DEVELOPMENT CORE TEAM (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`.

SAGHAEI, A., MEHRJOO, M. & AMIRI, A. (2009). A cusum-based method for monitoring simple linear profiles. *The International Journal of Advanced Manufacturing Technology* 45 1252–1260.

SHEWHART, W. A. (1939). *Statistical Method From the Viewpoint of Quality Control*. New York, NY: Dover Publications.

SHU, L., TSUNG, F. & TSUI, K. L. (2004). Run length performance of regression control charts with estimated parameters. *Journal of Quality Technology* 36 280–291.

SPARKS, R. S. (2000). CUSUM charts for signaling varying location shifts. *Journal of Quality Technology* 32 157–171.

STOVER, F. S. & BRILL, R. V. (1998). Statistical quality control applied to ion chromatography calibrations. *Journal of Chromatography A* 804 37–43.

SULLIVAN, J. H. (2002). Detection of multiple change points from clustering individual observations. *Journal of Quality Technology* 34 371–383.

SULLIVAN, J. H. & JONES, A. L. (2002). A self-starting control chart for multivariate individual observations. *Technometrics* 44 24–33.

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

WILLIAMS, W. H., WOODALL, W. H. & BIRCH, J. B. (2007). Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International* 23 925–941.

WOODALL, W. H. (2007). Current research on profile monitoring. *Produção* 17 420–425.

WOODALL, W. H., SPITZNER, D. J., MONTGOMERY, D. C. & GUPTA, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology* 36 309–320.

ZEISSET, M. S. (2008). *Wavelet methods in quality engineering: statistical process monitoring and experimentation for profile responses.* Ph.D. thesis, Florida State University, College of Engineering.

ZHANG, J., LI, Z. & WANG, Z. (2009). Control chart based on likelihood ratio for monitoring linear profiles. *Computational Statistics and Data Analysis* 53 1440–1448.

ZI, X., ZOU, C. & TSUNG, F. (2011). A nonparametric general linear profile monitoring .

ZOU, C., NING, X. & TSUNG, F. (2010). Lasso-based multivariate linear profile monitoring. *Annals of Operation Research (to appear)* .

ZOU, C. & QIU, P. (2009). Multivariate statistical process control using LASSO. *Journal of American Statistical Association* 104 1586–1596.

ZOU, C., QIU, P. & HAWKINS, D. (2009). Nonparametric control chart for monitoring profiles using change point formulation and adaptive smoothing. *Statistica Sinica* 19 1337–1357.

ZOU, C., TSUNG, F. & WANG, Z. (2007a). Monitoring general linear profiles using multivariate Exponentially Weighted Moving Average schemes. *Technometrics* 49 395–408.

ZOU, C., TSUNG, F. & WANG, Z. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics* 50 512–526.

ZOU, C., ZHANG, Y. & WANG, Z. (2006). A control chart based on a change-point model for monitoring linear profiles. *IIE Transactions* 38 1093–1103.

ZOU, C., ZHOU, C., WANG, Z. & TSUNG, F. (2007b). A self-starting control chart for linear profiles. *Journal of Quality Technology* 39 364–375.

Enrico Tonini

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4174
e-mail: tonini@stat.unipd.it

Current Position

January 2010 – January 2013
PhD Student in Statistical Sciences, University of Padova.
Thesis title: General Profile Monitoring Through Nonparametric Techniques
Supervisor: Prof. Guido Masarotto
Co-supervisor: Prof. Giovanna Capizzi.

Research interests

- Self-starting control charts
- General and nonparametric profile monitoring

Education

October 2008 – October 2009

Master (*laurea specialistica*) degree in "Scienze Statistiche, Demografiche e Sociali" (Statistics for Demography and Social Sciences).

University of Padova, Faculty of Statistics

Title of dissertation: "Innovative Statistical Methods for the Construction of Composite Scales from Ordinal Variables with Application to Business and Management Data"

Supervisor: Prof. Fortunato Pesarin

Final mark: 110/110

October 2005 – July 2008

Bachelor degree (*laurea triennale*) in Statistica, Popolazione e Società (Statistics, Population and Society).

University of Padova, Faculty of Statistics

Title of dissertation: "Anatomia di una distribuzione di probabilità" ("Anatomy of a Probability Distribution")

Supervisor: Prof. Guido Masarotto

Final mark: 110/110 e lode.

Awards and Scholarship

May 2009 – December 2009

University of Padova, Dip. di Tecnica e Gestione dei Sistemi Industriali

Progetto FIRB 2005 conto F.S.2.11.03.09, voce A2

Title: *Metodi di aggregazione di variabili discrete e analisi di affidabilità: applicazioni su dati aziendali*

Supervisor: Prof. Roberto Filippini

Computer skills

- R: very good
- SAS: good
- SPSS: good
- STATA: moderate
- LaTeX: very good

Language skills

Italian: native; English: good; Spanish: moderate; Dutch: moderate.

Conference presentations

Tonini, E., (2012). Self-starting control charts for monitoring general profiles using wavelet transforms and nonparametric techniques. (invited) *Joint Meeting of y-BIS–International Young Business and Industrial Statisticians and jSPE–Young Portuguese Statisticians*, Lisbon, Portugal, 23–26 July, 2012.

References

Prof. Guido Masarotto
University of Padova
Department of Statistics
via Cesare Battisti 241–243
Phone: +39 049 827 4145
e-mail: masarotto@stat.unipd.it