Working Paper Series, N. 20, Dicembre 2007

Practical Design of Generalized Likelihood Ratio Control Charts for Autocorrelated Data

Giovanna Capizzi
Department of Statistical Sciences
University of Padua
Italy

Guido Masarotto
Department of Statistical Sciences
University of Padua
Italy

Abstract: Control charts based on Generalized Likelihood Ratio (GLR) tests are attractive from both a theoretical and practical point of view. In particular, in the case of an autocorrelated process, the GLR test uses the information contained in the time-varying response after a change and, as shown by Apley and Shi, is able to outperfom traditional control charts applied to residuals. In addition, a GLR chart provides estimates of the magnitude and the time of occurrence of the change. In this paper, we present a practical approach to the implementation of GLR charts for monitoring an autoregressive and moving average process assuming that only a Phase I sample is available. The proposed approach, based on automatic time series identification, estimates the GLR control limits via stochastic approximation using bootstrap resampling. Thus, it is able to take into account the uncertainty about the underlying model. A Monte Carlo study shows that our methodology can be used to design in a semi-automatic fashion a GLR chart with a prescribed rate of false alarms when as few as 50 Phase I observations are available. A real example is used to illustrate the designing procedure.

Keywords: Automatic modeling; Autoregressive moving average models; Bootstrap; Statistical process control; Stochastic approximation; Uncertainty modeling.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Giovanna Capizzi
tel: +39 049 827 4168
giovanna.capizzi@unipd.it

Practical Design of Generalized Likelihood Ratio Control Charts for Autocorrelated Data

Giovanna Capizzi
Department of Statistical Sciences
University of Padua
Italy

Guido Masarotto
Department of Statistical Sciences
University of Padua
Italy

Abstract: Control charts based on Generalized Likelihood Ratio (GLR) tests are attractive from both a theoretical and practical point of view. In particular, in the case of an autocorrelated process, the GLR test uses the information contained in the time-varying response after a change and, as shown by Apley and Shi, is able to outperfom traditional control charts applied to residuals. In addition, a GLR chart provides estimates of the magnitude and the time of occurrence of the change. In this paper, we present a practical approach to the implementation of GLR charts for monitoring an autoregressive and moving average process assuming that only a Phase I sample is available. The proposed approach, based on automatic time series identification, estimates the GLR control limits via stochastic approximation using bootstrap resampling. Thus, it is able to take into account the uncertainty about the underlying model. A Monte Carlo study shows that our methodology can be used to design in a semi-automatic fashion a GLR chart with a prescribed rate of false alarms when as few as 50 Phase I observations are available. A real example is used to illustrate the designing procedure.

Keywords: Automatic modeling; Autoregressive moving average models; Bootstrap; Statistical process control; Stochastic approximation; Uncertainty modeling.

1 Introduction

Control charts are an effective tool in Statistical Process Control (SPC) for monitoring the stability of a process over time. Their performance is typically discussed in terms of run-length, rl, i.e., the number of observations between out-of-control signals. The run-length properties of traditional control charts, designed for independent and identically distributed (i.i.d.) observations, are strongly affected by data autocorrelation, a common feature of many processes. Thus, a process control scheme should be designed taking autocorrelation into account. Two methods for dealing with the violation of the i.i.d. assumption are proposed in the SPC literature.

In the first approach, the control limits of the traditional control charts are adjusted to account for the serial correlation in the data. Then, the original observations are plotted in control charts with modified limits (Vasilopoulos and Stamboulis, 1978; Yashchin, 1993; Schimd and Schöne, 1997; Runger, 2002). The second approach fits a time series model to the data. If both the structure and model parameters have been identified correctly, then the residuals are uncorrelated. Hence, traditional control charts can be applied to the residuals (Alwan and Roberts, 1988; Harris and Ross, 1991; Montgomery and Mastrangelo, 1991; Wardell et al., 1994; Runger and Willemain, 1995; Runger et al., 1995; Lu and Reynolds, 1999a,b; Shu et al., 2002). Both methods require knowledge of an appropriate time series model. The first approach uses the fitted model to set up the control limits, the second one to compute the residuals. Most early studies have assumed perfect knowledge of the mathematical models describing process behavior. In practice, the model is estimated from baseline data. Recent research shows that the combined effect of the variability of the estimators and model misspecification can significantly deteriorate the properties of the chart's run length performance for the autocorrelated data case and, hence, large sample sizes are needed to ensure that time series parameters can be estimated accurately (Adams and Tzeng, 1998; Boyles, 2000; Kramer and Schmid, 2000; Apley, 2002; Apley and Lee, 2003; Testik, 2005; Jensen et al., 2006). Further, model identification requires some skill and experience in time series analysis and, in addition, the power detection of traditional control charts, like EWMA and CUSUM, applied to residuals can be disappointingly poor in some situations since they do not take into account the fault signature, i.e., the time varying mean of the residuals resulting from a mean shift in the original process.

A solution to the last problem can be obtained using a proper change point model. The presence of a special cause of variation can be manifested by changes in mean or variance of a time series and also by the changes in its stochastic behaviour. A change point model supposes that time series models both before and after the change point can be described by a family of probability density function. A likelihood ratio test approach is used for deciding between two hypothesis: the null hypotesis that no change occurred and the alternative hypothesis that a change occurred at some intermediate unknown time. When the parameters of the distribution are known before and after the change point, the standard algorithm for the changepoint detection is the CUSUM test (Page, 1955). When the parameters after the change point are unknown, CUSUM can be generalized in several ways, including the Generalized Likelihood Ratio (GLR) test (Willsky and Jones, 1976; Basseville and Nikiforov, 1993; Superville and Adams, 1994; Siegmund and Venkatraman, 1995; Lai, 1995; Apley and Shi, 1999; Chang and Fricker, 1999; Lai, 2001; Runger and Testik, 2003). The GLR algorithm can be used for change point detection, location estimation, or tracking. Further, in the presence of a sudden step shift in the autocorrelated process, methods such as the GLR have the appealing advantage of incorporating the information contained in the fault signature. Because of its general applicability and attractive theoretical properties, the GLR is now considered a powerful tool for change point detection in dynamic systems.

However, a number of drawbacks with GLR have also been pointed out (see Basseville and Nikiforov, 1993; Lai, 1995). The GLR method implies a double max-

imization of the log-likelihood for each possible time of change point in a block of n observations. For very large n this computation could become a burden and various modifications procedures have been proposed to reduce the computational complexity. A possible approach is to restrict the maximum searching to a window of the M most recent observations, obtaining the so called window-limited GLR. This restricted search of the maximum offers computational advantages with essentially no loss of information (Lai, 1995). However, practical suggestions for choosing a suitable value of M and the corresponding control limit are still lacking (Willsky and Jones, 1976; Basseville and Nikiforov, 1993; Lai, 1995).

In order to address many of the previous issues, this paper discusses a practical design of GLR charts for detecting changes in the process mean and in the innovation variance of an autoregressive moving average process. We consider two types of GLR charts for jointly monitoring the process mean and variance, omnibus GLRs which make use of a single monitoring statistic and a combination of two GLRs one for the mean and another for the innovation variance. Observe that though we restrict our attention to the detection of either a change in the process mean or an increase of the innovation variance, the resulting control schemes offer some level of protection also against changes in the autocorrelation pattern. Indeed a change in the dynamic structure manifests itself also as an increase of the residual variance.

The paper is structured as follows. Section 2 present GLR charts assuming no model uncertainty. Section 3 discusses the impact of the estimation errors. Evidence shows that this effect cannot be neglected if a given false alarms rate must be mantained. The suggested designing procedure is described in Section 4. Model fitting is performed using the automatic three stages procedure introduced by Hannan and Rissanen (1982). Then, a bootstrap method coupled with stochastic approximation is used to tune control limits able both to satisfy a given constraint on the false alarms rate and to take into account the model uncertainty. In Section 5, we present an illustrative example using a real time series. The main results of a Monte Carlo experiment are presented in Section 6. Some details on the implementation, whose programs are publicy available upon request from the authors, are given in Section 7.

2 GLR control charts without model uncertainty

Assume that the in-control observations y_t can be described by a stationary autoregressive moving average $ARMA(p, q)$ process

$$y_t = \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \epsilon_t - \sum_{j=0}^{q} \theta_j \epsilon_{t-j} \tag{1}$$

where $\mu = E[y_t]$ and ϵ_t are Gaussian linear innovations with zero mean and variance σ^2. We will assume that the serial correlation is stable but a persistent change, occurring at the unknown time instant τ, is able to affect the mean of the original observations and/or the innovation variance. Hence, before τ the vector of parameters is assumed to be constant and equal to a known value $\beta = (\mu, \sigma^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$,

while after τ it is equal to $(\mu+\delta, \nu^2\sigma^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$, where $\delta \neq 0$ and $\nu^2 > 1$ are the unknown shift sizes.

A sequential change-point detection algorithm consists of testing the null hypothesis, H_0, that there is no change up to the current time t versus the alternative composite hypothesis, H_1, that a mean and/or a variance shift has occurred at an unknown instant τ, $1 \leq \tau \leq t$. A suitable decision function is evaluated each time a new observation is available and sampling is stopped when the decision is taken in favour of the alternative hypothesis.

An adequate tool for this testing hypothesis is the log-likelihood ratio test which compares the logarithm of the likelihood ratio between the two probability density functions for H_0 and H_1. Since the change time and the values of δ and ν^2 after the change are unknown, a solution consists of replacing each of the unknown parameters by their maximum likelihood estimate. This double maximization, limited to a finite horizon of length M, results in a window-limited Generalized Likelihood Ratio test, $GLRW$, given by

$$\max_{t-M+1\leq\tau\leq t} \sup_{\delta,\nu^2\geq 1} \sum_{i=\tau}^{t} \log \frac{f_1(y_i|y_{i-1},\ldots,y_1)}{f_0(y_i|y_{i-1},\ldots,y_1)} \tag{2}$$

where $f_1(\cdot)$ is the conditional density of y_t when at time τ both mean and/or variance are subjected to a shift of size δ and ν^2, respectively; while $f_0(\cdot)$ is the conditional density under the hypothesis of no change. The log-likelihood ratio, in (2), can be readily written as a function of the standardized innovations a_t computed, under the hypothesis of no change, using a Kalman-like filter. By well known properties of the innovation sequence in the Gaussian case (Brockwell and Davies, 1996), the standardized innovations are independent and

$$a_t \sim \begin{cases} N(0,1) & \text{when } t < \tau \\ N(\delta\rho(t,\tau), \nu^2) & \text{when } t \geq \tau \end{cases} \tag{3}$$

where $\rho(t,\tau)$ is the fault signature that reflects a pattern, left at time t, in the innovations by a unit magnitude mean shift occurring at τ (see Basseville and Nikiforov, 1993). Thus, step shifts in the mean and variance result in a time-varying mean of the innovation sequence and in a step profile of the innovation variance.

For the change-point model (3), the omnibus GLRW statistics, $GLRW_{OS}$, is given by

$$GLRW_{OS}(t) = 2 \max_{t-M+1\leq\tau\leq t} \sup_{\delta,\nu^2\geq 1} l_1(\tau,\delta,\nu^2)$$

where

$$l_1(\tau,\delta,\nu^2) = \sum_{i=\tau}^{t} \left\{ a_i^2 - \left(\frac{a_i - \delta\rho(i,\tau)}{\nu^2} \right)^2 \right\} - (t-\tau+1)\log\nu^2.$$

Let

$$s_t^2(\tau,\delta) = \frac{1}{t-\tau+1} \sum_{i=\tau}^{t} (a_i - \delta\rho(i,\tau))^2$$

Conditionally on τ, the maximum likelihood estimate, m.l.e, of the change magnitudes are given by

$$\hat{\delta}_t(\tau) = \frac{\sum_{i=\tau}^{t} a_i \rho(i, t-\tau)}{\sum_{i=\tau}^{t} \rho^2(i, t-\tau)} \quad \text{and} \quad \hat{\nu}_t^2(\tau) = \max\left\{1, s_t^2(\tau, \hat{\delta}_t(\tau))\right\}. \tag{4}$$

Hence, the $GLRW_{OS}$ can be otained by

$$GLRW_{OS}(t) = 2 \max_{t-M+1 \le \tau \le t} l_1(\tau, \hat{\delta}_t(\tau), \hat{\nu}_t^2(\tau)) \tag{5}$$

If the control statistic (5) falls beyond a suitable control limit h_{OS}, it is assumed that the process is off target. Thus, the run length of the $GLRW_{OS}$ chart is defined by the following stopping rule

$$rl_{OS} = \min\{t \ge 1 : GLRW_{OS}(t) \ge h_{OS}\} \tag{6}$$

Once a signal is given at t, the m.l.e. of the true change time is the value $\hat{\tau}$ which maximizes $l_1(\tau, \hat{\delta}_t(\tau), \hat{\nu}_t^2(\tau))$, and the m.l.e. of the change magnitudes, which can be used to have an initial idea of the fault's type, are $\hat{\delta}_t(\hat{\tau})$ and $\hat{\nu}_t(\hat{\tau})$.

In the change point model (3), potential shifts in the mean and/or variance have been assumed to occur at the same unknown instant of time τ, as usually discussed in the literature concerning the GLR charts (Lai, 1995, 2001). In the following, we will refer to (3) as the "*synchronous*" change point model. However, as the associate editor pointed out, mean and variance shifts might also occur at different times τ_1 and τ_2, respectively. In this case the standardized innovations are distributed according to

$$a_t \sim \begin{cases} N(0,1) & \text{when } t < \min(\tau_1, \tau_2) \\ N\left(\delta\rho(t, \tau_1), 1 + (\nu^2 - 1) I(t, \tau_2)\right) & \text{when } t \ge \min(\tau_1, \tau_2) \end{cases} \tag{7}$$

Here, $I(t, \tau_2)$ yields 1 when $t \ge \tau_2$ and 0 otherwise.

An omnibus GLRW statistic for the "*asynchronous*" change-point model (7) can be obtained by

$$\max_{(\tau_1, \tau_2) \in A(t)} \sup_{\delta, \nu^2 \ge 1} 2\, l_2(\tau_1, \tau_2, \delta, \nu^2) \tag{8}$$

where

$$l_2(\tau_1, \tau_2, \delta, \nu^2) = \sum_{i=\min(\tau_1, \tau_2)}^{t} \left\{ a_i^2 - \left(\frac{a_i - \delta\rho(i, \tau_1)}{1 + (\nu^2 - 1) I(t, \tau_2)} \right)^2 \right\} - (t - \tau_2 + 1) \log \nu^2$$

and $A(t) = \{(\tau_1, \tau_2) : t - M + 1 \le \tau_1, \tau_2 \le t\} \cup \{(\tau_1, \tau_2) : t - M + 1 \le \tau_1 \le t, \tau_2 = \infty\} \cup \{(\tau_1, \tau_2) : t - M + 1 \le \tau_2 \le t, \tau_1 = \infty\}$.

Given the values of τ_1 and τ_2, the m.l.e. of the change magnitudes, here denoted by $\hat{\hat{\delta}}$ and $\hat{\hat{\nu}}^2$, satisfy the following equations

$$\begin{cases} \hat{\hat{\delta}} = d_t(\tau_1, \tau_2, \hat{\hat{\nu}}^2) \\ \hat{\hat{\nu}}^2 = \max\left(1, g_t^2(\tau_1, \tau_2, \hat{\hat{\delta}})\right) \end{cases} \tag{9}$$

where

$$d_t(\tau_1,\tau_2,\nu^2)=\begin{cases}\hat{\delta}_t(\tau_1) & \text{when } \tau_1\geq\tau_2 \text{ or } t<\tau_2\\ \dfrac{\nu^2\sum_{i=\tau_1}^{\tau_2-1}a_i\rho(i,\tau_1)+\sum_{i=\tau_2}^{t}a_i\rho(i,\tau_1)}{\nu^2\sum_{i=\tau_1}^{\tau_2-1}\rho^2(i,\tau_1)+\sum_{i=\tau_2}^{t}\rho^2(i,\tau_1)} & \text{otherwise}\end{cases}$$

and

$$g_t^2(\tau_1,\tau_2,\delta)=\begin{cases}1 & \text{when } t<\tau_2\\ \sum_{i=\tau_2}^{t}(a_i-\delta\rho(i,\tau_1))^2/(t-\tau_2+1) & \text{otherwise.}\end{cases}$$

The system of equations (9) has no closed form solution when $\tau_1<\tau_2$. However, to avoid an iterative approach for solving the maximization problem, we suggest to use the following approximate statistics

$$GLRW_{OA}(t)=2\max_{(\tau_1,\tau_2)\in A(t)}\ l_2(\tau_1,\tau_2,\tilde{\delta}_t(\tau_1,\tau_2),\tilde{\nu}_t^2(\tau_1,\tau_2)) \tag{10}$$

where

$$\tilde{\delta}_t(\tau_1,\tau_2)=d_t(\tau_1,\tau_2,\tilde{\nu}_t^2(\tau_1,\tau_2)) \quad\text{and}\quad \tilde{\nu}_t^2(\tau_1,\tau_2)=\max\left\{1,g_t(\tau_1,\tau_2,\hat{\delta}_t(\tau_1)\right\} \tag{11}$$

with $\hat{\delta}_t(\tau_1)$ given by equation (4). Observe that expressions (11) are exact m.l.e. of the change magnitudes when $\tau_1\geq\tau_2$ or $\tau_2=\infty$, and approximate m.l.e. when $\tau_1<\tau_2$. The stopping rule for the asynchronous GLR test is given by

$$rl_{OA}=\min\{t\geq 1: GLRW_{OA}(t)\geq h_{OA}\}. \tag{12}$$

It is also straightforward to show that, for both the change-point models (3) and (7), the two GLRW control statistics for monitoring the process mean and the variance are given by

$$GLRW_{\delta}(t)=\max_{t-M+1\leq\tau\leq M-1}\ \sup_{\delta}\ l_1(\tau,\delta,1)=2\max_{t-M+1\leq\tau\leq t}l_1(\tau,\hat{\delta}_t(\tau),1) \tag{13}$$

and

$$GLRW_{\nu}(t)=\max_{t-M+1\leq\tau\leq M-1}\ \sup_{\nu^2\geq 1}\ l_1(\tau,0,\nu^2)=2\max_{t-M+1\leq\tau\leq t}l_1(\tau,0,\hat{\hat{\nu}}_t^2(\tau)) \tag{14}$$

respectively, with $\hat{\hat{\nu}}_t^2(\tau)=\max\left(1,s_t^2(\tau,0)\right)$. A signal is given if either of the two control statistics (13) and (14) exceeds the corresponding decision limits, say h_δ and h_ν. Thus, the stopping rule for the combined GLR algorithm, GLR_c, can be written as

$$rl_c=\min\{t\geq 1: GLRW_{\delta}(t)\geq h_\delta \ \text{ and/or } \ GLRW_{\nu}(t)\geq h_\nu\} \tag{15}$$

When the time series model is completely known a priori, the in-control run length distribution is completely known. Thus, for a given value of M, a suitable control limit h can be chosen so that a constraint on the frequency of false alarms is satisfied. Note that, in the following, when we speak of "false alarm rate" we are generically referring to a suitable measure of the incidence of false alarms, such as the probability of a false signal in a given time interval and the in control average run length.

3 Effects of model uncertainty

In the previous section, a perfect knowledge of the mathematical model describing process dynamics has been assumed. In real applications, the structure of the serial dependence is rarely known and a suitable time-series model is estimated from an in-control sample of n observations. Thus, the control chart is applied to the estimated residuals $a_t(\hat{\beta}_n)$, $\hat{\beta}_n$ being an estimate of the true parameters vector β. Due to the differences between $\hat{\beta}_n$ and β, the forecast residuals, $a_t(\hat{\beta}_n)$, are neither independent nor indentically distributed and the distribution function of the run length is not equal to that obtained in the known parameter case, at least when n is finite. A *naive* designing procedure consists of evaluating the critical values of the control chart under the assumption that the estimated model will exactly match the true model, i.e. assuming $\hat{\beta}_n = \beta$. Unfortunately, previous research concerning a variety of scenarios and monitoring procedures recognizes that model misspecification, or even small errors in the estimated parameters, lead to control chart performances that are often very different than those expected in the known parameters case (see Jensen et al., 2006). In particular, Apley and Shi (1999) show that the GLR chart, designed for mean shift detection in autocorrelated process, may also be adversely affected by modeling errors. In order to investigate how and to what extent modeling errors may affect a GLR chart, designed for jointly control both the mean and variance of autocorrelated data, we performed an extensive Monte Carlo study. Here, we report results for the following time series models:

$$M1.\quad y_t = \epsilon_t + 0.85\epsilon_{t-1}, \qquad M2.\quad y_t = 1.13y_{t-1} - 0.64y_{t-2} + \epsilon_t + 0.90\epsilon_{t-1},$$
$$M3.\quad y_t = 0.8y_{t-1} + \epsilon_t, \qquad M4.\quad y_t = 0.6y_{t-1} - 0.8y_{t-2} + 0.4y_{t-3} + \epsilon_t,$$

where $\epsilon_t \sim N(0, \sigma^2)$, with σ^2 such that that process variance is equale to one. The four ARMA models M1-M4 have been selected as their fault signatures are able to cover a wide variety of dynamics pattern (see Figure 1).

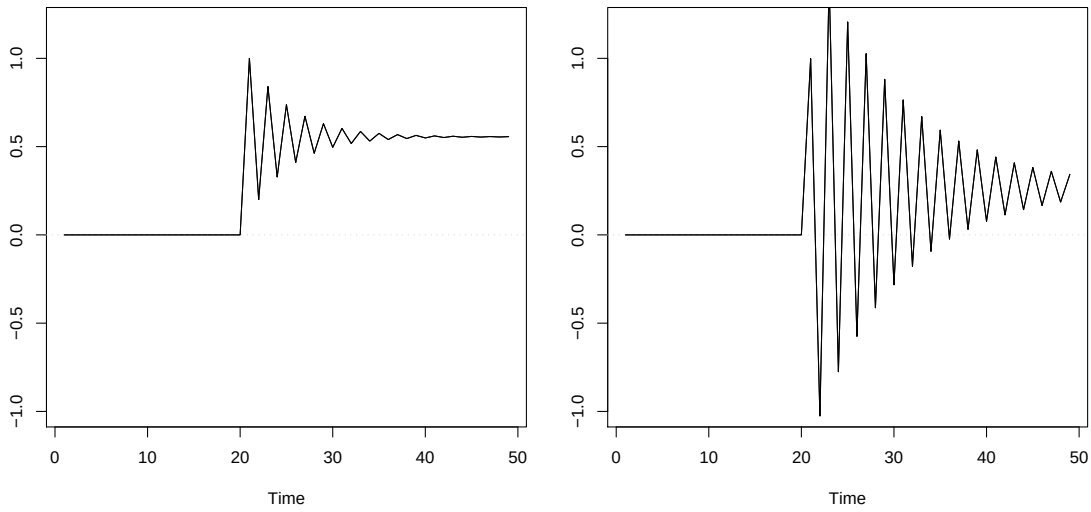
Three different cases of model knowledge are considered:

A. the underlying time series model is known *a priori*;
B. the model order is specified, but process parameters are unknown. The unknown parameters must be estimated from an in-control reference sample of size n. In particular, we make use of the exact maximum likelihood estimation method;
C. both the model order and time series parameters are unknown. Here, we apply the three stage procedure, suggested by Hannan and Rissanen (1982), modifying the second step as proposed by Kavalieris (1991). More details about the automatic order and parameter identification will be given in Section 4.

To evaluate the performance of the $GLRW$ charts with estimated parameters, we make use of both the conditional and marginal distribution of the in-control run length. The conditional run-length distribution is the probability mass function conditioned to the observed estimates while the marginal distribution is the conditional run length distribution averaged over the entire range of the estimates obtained from an in-control reference sample of the same size (Jones et al., 2001; Jones, 2002; Jensen et al., 2006).

Table 1 shows the average, the standard deviation and various lower and upper quantiles of the marginal in-control run length distribution for a $GLRW_{OS}$, with

$M1 : y_t = \epsilon_t + 0.8\epsilon_{t-1}$ $M2 : y_t = 1.13y_{t-1} - 0.64y_{t-2} + \epsilon_t + 0.9\epsilon_{t-1}$



$M3 : y_t = 0.8y_{t-1} + \epsilon_t$ $M4 : y_t = 0.6y_{t-1} - 0.8y_{t-2} + 0.4y_{t-3} + \epsilon_t$

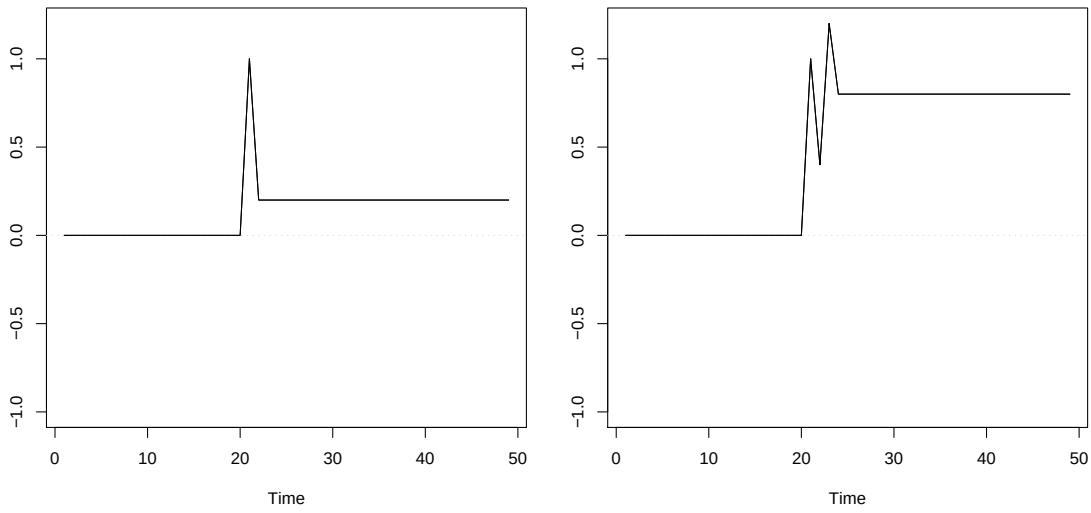


Figure 1: Fault signature of four ARMA models for a step mean shift at time $\tau = 21$.

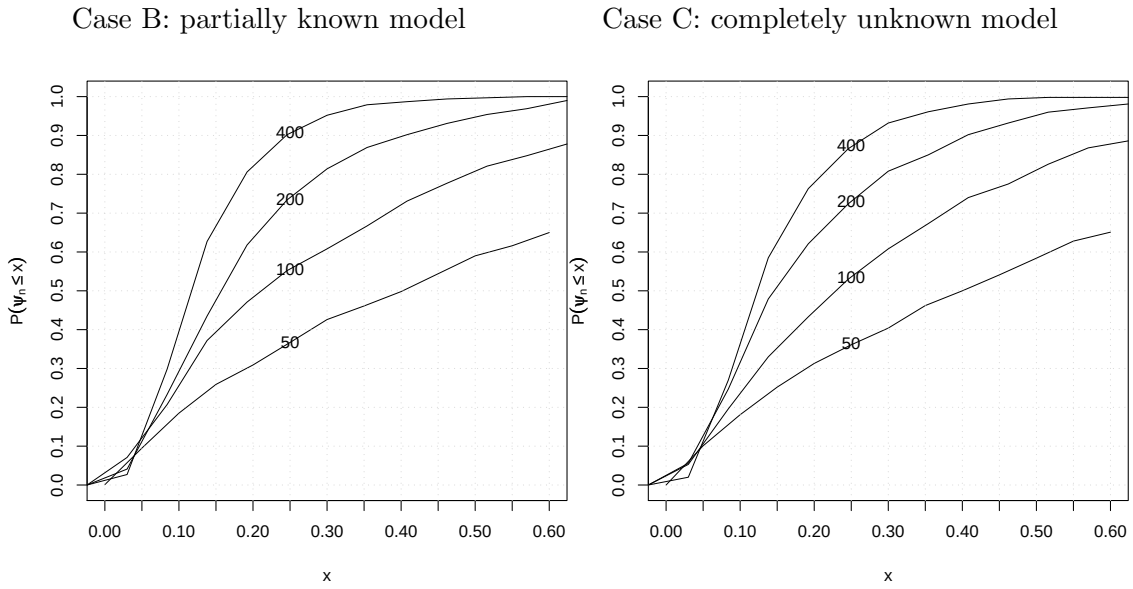


Figure 2: Estimated distribution of the conditional probability to give a false alarm within $t = 100$, $\psi_n = P(rl \leq 100|\hat{\beta}_n)$, when a omnibus $GLRW_{OS}$ chart ($M = 20$, $h_{OS} = 13.62199$), based on estimated parameters, is used to monitor data generated by model M3. Numbers on the curves give the sample sizes used to estimate the model.

$M = 20$, designed to give an in-control ARL roughly equal to 1000 when the model is known. Under the assumption that there are no modeling errors, case A, the control limits provide the desired ARL values and the standard deviation of the run length distribution is almost as large as the average run length. With model uncertainty, cases B and C, the estimated run length distribution appears to be left-shifted. That is, the main bulk of the run lengths takes relatively low values, but there is evidence of a long tail of high values. The occurrence of a relatively large percent of earlier false alarms, together with a few extremely long runs, results in a shift of the lower quantiles towards smaller values and in a standard deviation of the run length much larger than the ARL, especially with small sample sizes. Observe that, for $n > 100$, the performance of the estimated run-length distribution in case C is similar to that of case B. Thus, errors in parameter estimation seem to have a stronger effect on the run-length performance than possible model misspecification. On the other hand, the extent to which the GLR chart is sensitive to false alarms when parameters are estimated seems to depend on the particular generating model. Using model M1, for instance, the median and upper quantiles of the estimated run length distribution are larger and smaller, respectively, than the corresponding quantiles for model M4.

Figure 2 illustrates the distribution of the conditional probability of obtaining a false signal within $t = 100$, i.e. $P(rl \leq 100|\hat{\beta}_n)$, for model $M3$ in the presence of

Table 1: In control performance of $GLRW_{OS}$ control charts ($M = 20$) with an in-control ARL approximately equal to 1000 when the model is known.

	A: known model	B: partially known model				C: completely unknown model			
		50	100	200	400	50	100	200	400
				Model M1 ($h_{OS} = 13.92588$)					
ARL	1002.1	977.2	918.9	943.7	971.3	836.7	854.8	925.5	959.6
SDRL	991.1	7102.1	2217.2	1478.6	1211.4	8346.0	2164.1	1515.8	1224.7
$Q_{0.01}$	15	6	8	11	13	4	7	10	13
$Q_{0.25}$	294	63	119	186	231	44	97	171	225
$Q_{0.50}$	700	188	342	482	578	140	294	457	561
$Q_{0.75}$	1392	604	906	1121	1247	469	820	1081	1226
$Q_{0.99}$	4567	12230	8673	6860	5772	10833	8779	6878	5778
				Model M2 ($h_{OS} = 13.59065$)					
ARL	999.5	1227.0	958.4	949.9	953.5	814.3	859.8	907.8	959.9
SDRL	984.5	21695.4	3668.9	1788.4	1315.5	6855.4	2821.4	1623	1321.6
$Q_{0.01}$	13	4	7	10	12	4	7	9	11
$Q_{0.25}$	285	59	106	160	207	39	90	151	205
$Q_{0.50}$	686	180	301	429	530	118	260	410	532
$Q_{0.75}$	1374	560	823	1053	1184	382	720	1006	1194
$Q_{0.99}$	4563	14977	10419	7719	6216	10904	9693	7517	6272
				Model M3 ($h_{OS} = 13.62199$)					
ARL	997.2	1159.5	952.8	927.0	944.2	1055.5	918.3	910.5	932.3
SDRL	991.4	11620	3439.9	1834.5	1352.3	9256.1	3667.8	1775.8	1327.2
$Q_{0.01}$	16	5	8	10	13	4	7	10	12
$Q_{0.25}$	294	42	85	143	199	34	79	138	194
$Q_{0.50}$	696	129	254	391	510	109	237	385	498
$Q_{0.75}$	1377	460	766	998	1158	401	721	983	1140
$Q_{0.99}$	4598	16803	10905	7899	6387	15324	10594	7809	6370
				Model M4 ($h_{OS} = 13.83589$)					
ARL	998.7	892.4	878.7	918.3	945.1	663.7	788.1	880.1	938.0
SDRL	992.1	7261.5	2788.7	1690.2	1311.8	6710.4	2616.1	1657.9	1288.4
$Q_{0.01}$	14	4	6	9	12	3	6	9	11
$Q_{0.25}$	293	44	92	155	207	28	75	141	204
$Q_{0.50}$	693	135	267	417	527	83	221	384	518
$Q_{0.75}$	1383	439	754	1013	1174	278	652	956	1165
$Q_{0.99}$	4548	11779	9759	7628	6087	9369	9046	7577	6102

Note. 50, 100, 200, 400 are the sample sizes used to fit the model. *ARL*, *SDRL* and Q_p denote the average, the standard deviation and the p-quantile of the marginal in-control run-length distribution (estimated using 100000 Monte Carlo replications).

Table 2: Percent increase in the marginal false-alarm probability of the $GLRW_{OS}$ chart with parameter estimated.

	Sample size											
Model	50	100	200	400	600	800	1000	1200	1400	1600	1800	2000
	Case B: partially known model											
M1	290.8	143.8	61.3	30.0	20.6	16.0	10.8	10.2	8.9	8.8	8.5	8.1
M2	277.7	150.3	75.5	36.2	24.3	14.5	12.3	9.4	6.8	6.3	4.2	3.7
M3	395.5	218.2	107.1	50.8	35.2	25.2	20.8	18.2	14.2	13.1	12.4	10.5
M4	369.7	192.6	90.7	41.7	28.3	20.0	18.2	15.5	11.9	10.3	10.0	8.4
	Case C: completely unknown model											
M1	375.5	186.0	76.3	33.2	22.2	18.2	13.7	10.7	9.9	8.7	8.6	6.9
M2	382.9	184.4	84.5	38.9	22.8	17.6	13.9	11.1	9.4	7.4	6.2	6.4
M3	441.1	235.0	113.0	54.8	35.9	25.9	18.7	17.7	16.2	13.4	12.5	12.3
M4	497.4	241.6	106.9	46.3	29.0	23.3	19.0	14.8	13.4	12.2	11.8	8.7

Note. Probabilities have been estimated using 100000 Monte Carlo replications.

model uncertainty. When the true model is known, the probability of an incorrect signal before the 100th observations is roughly equal to 0.09. Observe that, even for $n = 400$, the probability that a false alarm is triggered can be 100 percent larger than the nominal value with probability about equal to 0.20.

Table 2 shows, for a range of sample sizes, the percent increase in false alarm rates at $t = 100$ for $GLRW_{OS}$ charts using estimates over those with known parameters. Quesenberry (1993) and Jones et al. (2001) pointed out that for a chart with estimated parameters to perform like one with known parameters a negligible increase in the false alarm rate would be required. Considering the case of independent data, the authors use tables like Table 2 to determine the minimum sample size necessary so that the Shewhart and EWMA control charts with estimated parameters performs as the corresponding charts with known parameters. In the present case, the decrease with n seems to depend upon the particular ARMA model. Hence, it is difficult to give precise guidelines on the necessary sample size. However, observe that if an increase in the false alarm rate of no more than 10% is considered acceptable, as discussed in (Jones et al., 2001), more than 2000 observations are required for a $GLRW_{OS}$ to perform like one with known parameters for the model M3.

Note that similar results have been obtained for other time series models and for different types of GLR charts. In addition, results for other values of the window size M, which have been omitted here for lack of space, show a very similar pattern but also a stronger impact of the estimation errors when a larger value of M is used. This is consistent with analogous studies, in the framework of independent data, showing that the effect of parameters estimation, on both the in-control and out-of-control run length performance, is much larger when charts with a longer memory are used, e.g. Jones et al. (2001). A related example is given in Table 3 listing, for a variety

Table 3: Marginal probabilities of detection within 20 timesteps of a change point occuring at $t = 100$ of $GLRW_{OS}$ charts used to monitor data generated by model M3.

	Out of control situations							
	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 2$	$\delta = 0$	$\delta = 0$	$\delta = 0$	$\delta = 0$
M	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1.25$	$\nu = 1.5$	$\nu = 2$	$\nu = 2.5$
Case A: known model								
10	0.041	0.137	0.401	0.748	0.198	0.561	0.950	0.995
20	0.047	0.181	0.516	0.855	0.203	0.584	0.960	0.996
30	0.050	0.186	0.516	0.854	0.201	0.578	0.959	0.996
Case B: partially known model ($n = 50$)								
10	0.028	0.077	0.196	0.420	0.071	0.237	0.721	0.946
20	0.028	0.085	0.229	0.471	0.062	0.215	0.722	0.948
30	0.027	0.082	0.206	0.423	0.056	0.190	0.678	0.935
Case B: partially known model ($n = 100$)								
10	0.033	0.102	0.293	0.596	0.128	0.396	0.871	0.982
20	0.036	0.122	0.351	0.676	0.117	0.383	0.876	0.983
30	0.033	0.116	0.326	0.642	0.106	0.358	0.860	0.982
Case C: completely unknown model ($n = 50$)								
10	0.026	0.063	0.167	0.356	0.054	0.192	0.659	0.925
20	0.026	0.072	0.191	0.408	0.048	0.170	0.657	0.928
30	0.025	0.066	0.170	0.364	0.044	0.150	0.614	0.911
Case C: completely unknown model ($n = 100$)								
10	0.033	0.099	0.279	0.575	0.116	0.371	0.861	0.981
20	0.035	0.122	0.331	0.654	0.106	0.368	0.869	0.985
30	0.034	0.117	0.308	0.626	0.098	0.349	0.855	0.982

Note. Critical values of all schemes have been chosen so that the in-control probability to signal within $t = 100$ is marginally equal to 0.1. Probabilities have been estimated using 100000 Monte Carlo replications.

of change-point scenarios and for different values of the window size, the marginal detection probabilities of three $GLRW_{OS}$ charts designed for monitoring residuals from the $AR(1)$ model $M3$. To make this comparison possible, the critical value h_{OS} of the compared schemes have been chosen so that the probability of a false alarm before $t = 100$, i.e. $P(rl \leq 100)$, is marginally equal to 0.1. Observe that there is a large deterioration in the chart performance due to the estimation errors which, however, is relatively more pronounced for larger values of M. Consequently, when parameters are estimated, the GLR chart based on $M = 20$ performs sligtly better than that based on $M = 30$, notwithstanding no practical difference can be found in the marginal detection probabilities of detection when the model is perfectly known. However, in both the case of known and unknown model, differences between the probabilities of detection, computed for different values of M, cannot be considered

remarkable. Furthermore, results shown in Table 3 confirm that, at least for $n = 100$, the unknowing of model parameters have a larger impact on the chart performance than the unknowing of model order.

We have seen that the GLR test takes explicitely into account the fault signature left in the residuals by a shift in the mean of the original observations. For this reason, it is able to outperform standard monitoring schemes, such as EWMA and CUSUM charts, suffering from a problem called "forecast recovery" problem (Superville and Adams, 1994). In the presence of modelling errors, and when large sample sizes are not available, it might be worthwhile to investigate whether the GLR chart based on the estimated fault signature still outperforms the standard monitoring algorithms. Apley and Shi (1999) have already presented comparisons between GLR and residual-based Shewhart and CUSUM tests for step mean shifts. Authors have shown that, when model uncertainty is present, the GLR continues to outperform these two standard control schemes. Here, we present the comparison between the combined GLR, defined by the stopping rule (15), and the combined EWMA (Lu and Reynolds, 1999a), whose stopping rule is defined by

$$rl_E = \min\{t \geq 1 : u_{\delta,t} \geq h_{E,\delta} \text{ and/or } u_{\nu,t} \geq h_{E,\nu}\},$$

where

$$u_{\delta,t} = (1-\lambda)u_{\delta,t-1} + \lambda a_t \text{ and } u_{\nu,t} = (1-\lambda)\max(1, u_{\nu,t-1}) + \lambda a_t^2$$

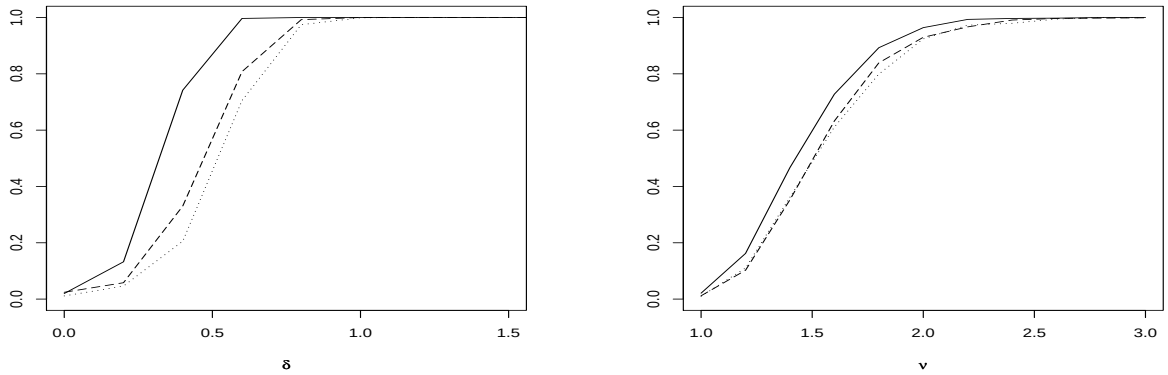
with $u_{\delta,0} = 0, u_{\nu,0} = 1$. As criterion for comparison we use the probability of detecting a shift within 20 timesteps after its occurrence for charts designed to have marginally the same probability to give a false alarm before $t = 100$. The detection probabilities are plotted in the Figure 3 as a function of δ and ν. Observe that when the model is perfectly known, case A, the combined GLR performance is better than that of the combined EWMA over all the range of δ and ν values. When model uncertainty is considered, cases B and C, and the combined tests are designed using a small reference sample, even if the power detection of both control schemes seems to be strongly affected by modeling errors, the combined GLR clearly continues to outperform the combined EWMAs. Similar results have been obtained using other time series models and different versions of the GLR chart.

Results of this section show that, although the GLR chart seems to be more efficient than other control charts even in the case of modeling errors, the effect of uncertainty on its performance cannot be neglected unless the number n of observations gathered in Phase I is extremely large. For this reason, in the next section we introduce an alternative designing method that, taking into account the variability of parameters estimation, leads to charts able to achieve a prescribed rate of false alarm with a reasonable number of observations.

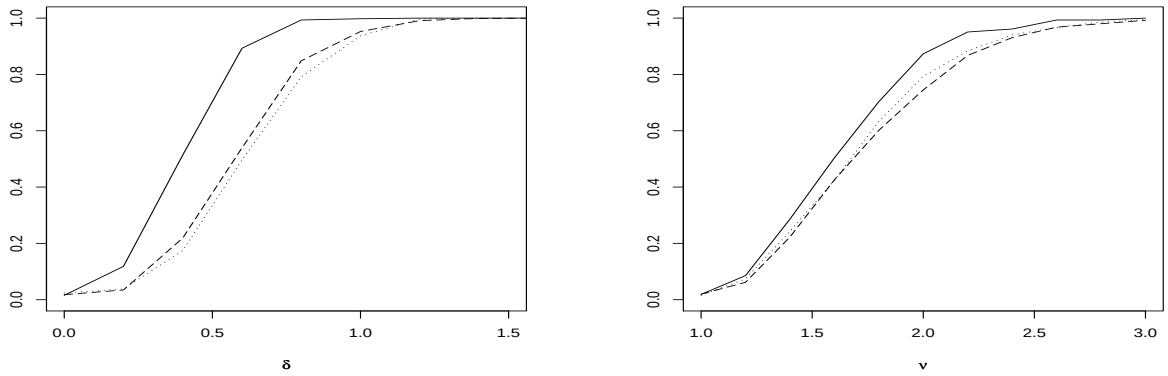
4 A bootstrap design of GLR control charts

Let $H_n(\cdot, \beta, \tau, \delta, \nu^2, h, M)$ be the marginal run length distribution when n in-control Phase I observations, $y_1, \ldots, y_n$, are used to fit the time series model. Here, the decision interval h is h_{OS} or h_{OA}, in the case of the omnibus charts (5) and (10),

Case A: known model



Case B: partially known model, $n = 50$



Case C: completely known model, $n = 50$

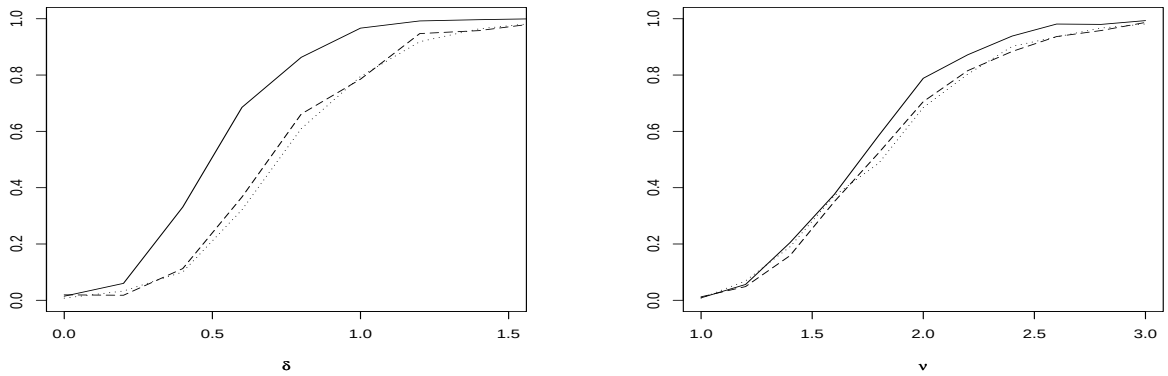


Figure 3: Probability of detecting step mean and variance shifts within the 20^{th} timestep after their occurrence at $t = 100$, for a combined GLR ($M = 20$, solid curve), and for two combined EWMAs, ($\lambda = 0.1$ and $\lambda = 0.5$, dashed and dotted curves, respectively), designed to monitor residuals from M2. For all charts, the control limits are chosen so that $P(rl \leq 100) = 0.1$, marginally.

respectively, (h_δ, h_ν) in the case of the two single charts (13) and (14). The instant τ is either the unknown single change time for the model (3) or the unknown vector (τ_1, τ_2) for the asynchronous model (7).

The designing procedure of a window limited GLR requires determining adequate values of both the control limit h and the window size M. In particular, for a given value of M, the critical value h is usually determined so that a given constraint on the in-control run-length distribution can be satisfied. However, when β is unknown, the exact computation of $H_n(\cdot, \beta, \tau, \delta, \nu^2, h, M)$ is not possible and the GLR design needs to be based on an estimated run-length distribution.

4.1 Estimation of the run-length distribution

Suppose that the order and coefficients of the $ARMA(p, q)$ process are unknown and have to be estimated. Automatic model identification and a bootstrap-based approach can be used to estimate the GLR run length distribution, $H_n(\cdot; \beta, \tau, \delta, \nu^2, h, M)$, taking into account the errors due to model misspecification and parameter estimation, (Bühlmann, 2002).

The method proceeds as follows:

i) Estimate the order and parameters of the generating model using the three stage algorithm developed by Hannan and Rissanen (1982). Since the original procedure may overestimate the model orders, the second step is modified following the Kavalieris (1991) proposal that guarantees consistency with respect to the order selection criterion.

   Briefly, the algorithm can be summarized as follows. In the first stage, a high order autoregressive model is selected by the AIC criterion, with a maximum order fixed at $10 \log_{10}(n)$. Then, the selected model is used to estimate the error terms ϵ_t. In the second stage, $ARMA(\tilde{p}, \tilde{q})$ processes, with $\tilde{p} < p_{\max}$ and $\tilde{q} < q_{\max}$, are fitted by regressing y_t on its $\tilde{p}$ past values and on the $\tilde{q}$ lagged values of error terms obtained at the previous stage (in all our examples, we use in particular $p_{\max} = q_{\max} = 5$). Let $(\hat{p}, \hat{q})$ be the best order according to the BIC criterion. The third stage consists of obtaining the m.l.e. $\hat{\beta}_n = (\hat{\mu}, \hat{\sigma^2}, \hat{\phi}_1, \ldots, \hat{\phi}_{\hat{p}}, \hat{\theta}_1, \ldots, \hat{\theta}_{\hat{q}})$.

ii) Generate pseudo-data y_t^* according to

$$y_t^* = \hat{\mu} + \sum_{j=1}^{\hat{p}} \hat{\phi}_j (y_{t-j}^* - \hat{\mu}) + \epsilon_t^* - \sum_{j=1}^{\hat{q}} \hat{\theta}_j \epsilon_{t-j}^*, \quad t = 1, 2, \ldots \tag{16}$$

   with $\epsilon_t^* \sim N(0, \hat{\sigma}^2)$.

   Equation (16) can also be used to generate observations under the out-of control situation. For example, under the synchronous model (3), from the instant of time $n + \tau$ a constant δ may be added to the pseudo-data y_t^*, when the change affects the process mean, and/or ϵ_t^* may be generated from $N(0, \nu^2 \hat{\sigma}^2)$ under the hypothesis of a shift in the innovation variance.

iii) Use $(y_1^*, \ldots, y_n^*)$, the first n observations from (16), to select the model order $(\hat{p}_n^*, \hat{q}_n^*)$. Then, compute the m.l.e. $\hat{\beta}_n^* = (\hat{\mu}^*, \hat{\sigma}^{2*}, \hat{\phi}_1^*, \ldots, \hat{\phi}_{\hat{p}_n^*}^*, \hat{\theta}_1^*, \ldots, \hat{\theta}_{\hat{q}_n^*}^*)$, as in step i).

iv) Apply the control chart to the sequence $a^*_{n+1}(\hat{\beta}^*_n), a^*_{n+2}(\hat{\beta}^*_n), \ldots$, where $a^*_t(\hat{\beta}^*_n)$ are the residuals computed for y^*_t. Define rl^* equal to $T - n$, where T is the first time at which the control chart gives an out-of-control signal.

Observe that step iv) assumes that data to be monitored are a continuation of the original trajectory. However, the procedure can be also adapted when the residual sequence corresponds to an independent trajectory generated by the same process.

v) Repeat steps ii)-iv) a relatively large number of times and use the empirical distribution of rl^* to approximate the unknown distribution $H_n(\cdot; \beta, \tau, \delta, \nu^2, h, M)$.

Observe that, as the number of boostrap replications goes to infinity, the empirical distribution of the run-lengths rl^* tends to $H_n(\cdot; \hat{\beta}_n, \tau, \delta, \nu, h, M)$ that is, hence, used to estimate the unknown distribution function $H_n(\cdot; \beta, \tau, \delta, \nu, h, M)$.

4.2 Estimation of the control limit

One of the most common criterion for choosing a suitable control limit consists of identifying a constant h which provides, for a given value of M, a control scheme with a desired in control average run length, ARL_0, i.e. such that $E_0(rl) = ARL_0$, where $E_0(\cdot)$ indicates the expected value of the run length under the hypothesis of no change. However, when the run length distribution exhibits an asymmetric and heavy-tailed behavior, as is the case with estimated parameters, it is important to investigate a more robust measure than the ARL, at least when large sample sizes are not available and/or a very low frequency of false alarms to be achieved. Indeed, as we have shown in Section 3, the ARL may be unable to capture the increase in the incidence of false signals. A reasonable alternative might consist of determining the control limit h so that the probability of a false alarm before some specified value, N_0, is equal to a desired value p_0, i.e. solving for h

$$H_{0,n}(N_0; \beta, h, M) = P(rl \leq N_0) = p_0 \tag{17}$$

where $H_{0,n}(\cdot; \beta, h, M)$ is the marginal in-control run-length distribution. Observe that, when $p_0 = 0.50$, this approach corresponds to the median run length criterion discussed by Gan (1993, 1994) and Lee and Khoo (2006). A similar constraint on the GLR run-length distribution has also been proposed by Lai (2001) as a more tractable and more appropriate criterion than the ARL when the observations are not independent. This kind of approach also looks appealing because the control chart performance can be improved updating the control limit as more data become available. In this case, N_0 can be set equal to the number of observations up to the next updating and p_0 equal to an acceptable rate of false alarms.

In the presence of model uncertainty, the bootstrap estimate of the probability of false alarms before N_0, i.e. $H_{0,n}(N_0; \hat{\beta}_n, h, M)$, may be used for determining the control limit able to achieve desired performance. In this case, a suitable value of the control limit for both the omnibus charts, (5) and (10), can be obtained by solving for h the following equation

$$H_{0,n}(N_0; \hat{\beta}_n, h, M) = p_0 \tag{18}$$

The solution of (18) can readily be computed via stochastic approximation. In particular, we have obtained satisfactory results using the Polyak-Ruppert algorithm (Ruppert, 1991; Polyak and Juditsky, 1992) which is both efficient and simple to implement. Following this approximation algorithm the h estimate at the $s-$th step is

$$\hat{h}_s = \frac{1}{s} \sum_{i=s_0+1}^{s_0+s} h_i \tag{19}$$

where the h_i values are generated by the recursion

$$h_{i+1} = \max[0, h_i - Ai^{-\alpha} q_i], \quad i = 1, 2, \ldots \tag{20}$$

Here, $q_i = I(N_0, rl_i^*) - p_0$, where $I(\cdot, \cdot)$ is the indicator function defined in Section 2; rl_i^* are independent random variables, with distribution $H_{0,n}(\cdot; \hat{\beta}_n, h_i, M)$ simulated via steps i)-iv) described in the subsection 4.1; h_1 is an initial guess of h, while A, s_0, s and α are suitable constants. The max operator in (20) is here used to ensure that the h estimates are greater than zero at each step. In our experiments, however, the h estimates happened to be negative only at the first steps of the iterative algorithm (20).

Theoretical results (Kushner and Yin, 2003) ensure that, for arbitrary values of constants $A > 0$, $s_0 \geq 0$, and $0.5 < \alpha < 1$, $H_{0,n}(N_0; \hat{\beta}_n, \hat{h}_s, M)$ converges almost surely to p_0, as s tends to infinity. In addition, observe that, for sufficiently large values of s, $H_{0,n}(N_0; \hat{\beta}_n, \hat{h}_s, M)$ is approximately normally distributed with mean p_0 and variance $p_0(1-p_0)/s$. This result may be used for suitably choosing the number of iterations. For example, if $p_0 = 0.1$ and $s = 10000$, the previous result ensures that $|H_{0,n}(N_0; \hat{\beta}_n, \hat{h}_s, M) - p_0|$ is less than 0.006 with probability approximately equal to 0.95.

The Polyak-Ruppert algorithm can be also used for designing the combined GLR chart, defined by the stopping rule (15). Since in the combined case two control limits, h_δ and h_ν, need to be determined, a new constraint must be added to that ensuring a desired joint false alarm rate. It seems reasonable, here, to balance the false alarm probability between the two charts, that is to impose the following constraint

$$P(rl_\delta < rl_\nu) = P(rl_\delta > rl_\nu) \tag{21}$$

where

$$rl_\delta = \min\{t > 1 : GLRW_\delta \geq h_\delta\}, \quad rl_\nu = \min\{t > 1 : GLRW_\nu \geq h_\nu\}. \tag{22}$$

The critical values (h_δ, h_ν) of the stopping rule (15), solution of the system of equations (18) and (21), can be computed by iterating the following recurrence relations

$$\begin{cases} h_{\delta,i+1} = \max[0, h_{\delta,i} - Ai^{-\alpha}(q_i - z_i)], & i = 1, 2, \ldots \\ h_{\nu,i+1} = \max[0, h_{\nu,i} - Ai^{-\alpha}(q_i + z_i)], & i = 1, 2, \ldots \end{cases} \tag{23}$$

with $q_i = I(N_0, rl_{c,i}^*) - p_0$ and $z_i = I(rl_{\nu,i}^*, rl_{\delta,i}^*) - I(rl_{\delta,i}^*, rl_{\nu,i}^*)$, and averaging the results as in (19). In each of the previous expressions the asterisk denotes the bootstrap replicates of the run lenghts. Let $\hat{h}_s$ denote the approximate bivariate

solution. It is straightforward to show that theoretical results used to suitably choosing the iteration number in (20) can also be applied to stop the recursive estimation (23). In particular, $H_{0,n}(N_0, \hat{\beta}_n, \hat{h}_s, M)$ is asymptotically distributed as in the omnibus case.

Analogously, stochastic approximation can also be used to obtain critical values ensuring a prescribed in-control ARL value. In particular, we yield satisfactory results using either (20) or (23) with q_i replaced by $(rl_i^* - ARL_0)/ARL_0$. However, as discussed in Section 3, in the presence of modeling errors, the in-control run length distribution is very-right skewed. Hence, we strongly recommend that users evaluate more meaningful performance measures than the average run length. Incidentally, observe that the quantile criterion (18) also offers some computationally advantages regarding the iterative approximation algorithm. Indeed, when (18) is used, we always truncate at $t = N_0$ each bootstrap replicate of the run-length, rl^*. Thus, very long execution times, due to the occurrence of extremely large run lengths, can be avoided. A similar truncation can be introduced also for the ARL criterion, but it would correspond to a much higher value. Indeed, when the ARL criterion is used, we have found convenient to arrest the run-lengths simulation at the value $\bar{rl}_i$ such that

$$h_{i+1} - h_i \leq \xi, \;\; \text{i.e. } \bar{rl}_i \leq ARL_0 \left(1 + \frac{\xi i^\alpha}{A}\right)$$

where ξ is a constant which, in our experience, can be set equal to 2 or 3. In this way, we avoid that the sequence h_i wanders too much and also unuseful long simulation runs.

4.3 Choice of window length

As discussed in Section 3, choosing of window length is less critical than that of control limit. However, window M should be selected to guarantee a satisfactory out-of-control performance. We suggest, in particular, to evaluate the GLR power detection for a range of M values. To provide an appropriate basis for comparison, all charts shoul be designed to have a common marginal in-control run-length property. Following this approach, in our opinion, the final choice of the window length should be left to practitioners. Users can prefer a window size able to guarantee a satisfactory performance for a wide range of change magnitudes rather than have the best power detection just for one specified shift. Observe that the analysis of several values of M is computationally unexpensive since computation of both the control limit and probability of detection (or even other out-of-control performance measures) can be done simultaneously for a range of values of the window length. Indeed, once a time series model has been fitted to a Phase I sample, run-lengths of the GLR charts, based on the different values of M, can be computed from the same bootstrap sequence $y_1^*, y_2^*, \ldots$. Thus, the model needs to be re-fitted only once; in addition, given the nested structure of the GLR statistics, the computational burden, for different values of the window size, is roughly equivalent to that requested by the maximum value of M (see also Section 7).

Table 4: Bootstrap estimates of the detection power of $GLRW_{OS}$ charts for a range of values of M, δ and ν.

	Window size					
$P_{\delta,\nu}(r)$	5	10	15	20	25	30
			$p_0 = 0.1$			
$P_{1,1}(10)$	0.076	0.075	0.059	0.052	0.045	0.042
$P_{1,1}(20)$	0.101	0.099	0.102	0.105	0.097	0.090
$P_{2,1}(10)$	0.497	0.512	0.446	0.406	0.380	0.358
$P_{2,1}(20)$	0.536	0.550	0.555	0.568	0.545	0.525
$P_{0,1.5}(10)$	0.179	0.151	0.118	0.099	0.088	0.082
$P_{0,1.5}(20)$	0.335	0.306	0.289	0.270	0.249	0.235
$P_{0,3}(10)$	0.945	0.937	0.928	0.915	0.907	0.902
$P_{0,3}(20)$	0.997	0.996	0.998	0.998	0.997	0.997
			$p_0 = 0.2$			
$P_{1,1}(10)$	0.122	0.128	0.110	0.101	0.096	0.092
$P_{1,1}(20)$	0.158	0.174	0.177	0.186	0.180	0.172
$P_{2,1}(10)$	0.645	0.669	0.630	0.600	0.580	0.565
$P_{2,1}(20)$	0.690	0.710	0.736	0.753	0.735	0.723
$P_{0,1.5}(10)$	0.292	0.267	0.243	0.225	0.213	0.202
$P_{0,1.5}(20)$	0.504	0.484	0.478	0.469	0.453	0.437
$P_{0,3}(10)$	0.965	0.964	0.959	0.955	0.952	0.951
$P_{0,3}(20)$	0.999	1.000	0.999	0.999	0.999	1.000

5 An Example

The design procedure, given in the previous section, is illustrated here by applying a $GLRW_{OS}$ chart to a real data set. In particular, we consider Series A from Box et al. (1994) that consists of 197 concentration readings taken every two hours from a chemical process.

Phase I. The first 150 of these measurements are used to fit an $ARMA(p, q)$ model. The identification of an appropriate order and the estimate of the coefficients model is done by the automated procedure, described in Section 4, leading to the following $ARMA(1, 1)$ model

$$y_t = 16.975 + 0.930(y_{t-1} - 16.975) + \epsilon_t + 0.654\epsilon_{t-1}, \quad \hat{\sigma}^2 = 0.097 \tag{24}$$

Diagnostic analysis shows no presence of residual correlations (the Box-Liung test based on the first 20 lags is equal to 21.2814 with a p value of 0.3807) and no violation of the normality assumption (Shapiro-Wilk statistic=0.9915, with $p =$ 0.5062). Since outliers in the data might affect the model's adequacy, it is also necessary to check their presence and adjust the model parameters appropriately. However, using the outlier identification procedure described by Wei (2006), no

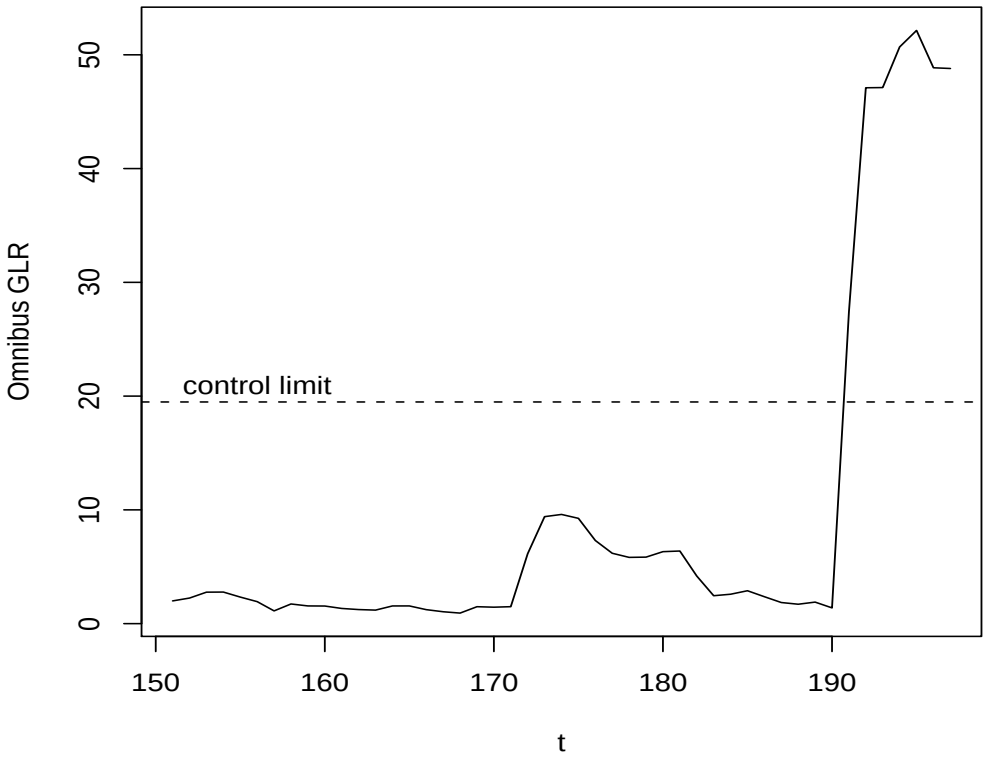


Figure 4: $GLRW_{OS}$ chart ($M = 10$, $h_{OS} = 19.48519$) applied to the Series A from Box et al. (1994)

significant additive outlier and/or level shift is identified. Thus, we assume that first 150 observations come from an in-control process.

Suppose that an appropriate constraint on the rate of false alarms is given by (17), with $N_0 = 100$ and $p_0 = 0.1$. In order to select an adequate $GLRW_{OS}$ chart we compute, for $M = 5, 10, 15, 20, 25, 30$, the bootstrap control limit h so that the probability of an incorrect signal within the 100-th observation is equal to 0.1. Then, we estimate the probabilities $P_{\delta,\nu}(r)$ of detecting a mean and/or a variance shift after a certain number, r, of timesteps have passed from a change point occurring at $t = 101$. Results are shown in Table 4 for a range of values of δ, ν and r. Note that the choice of the window size M slightly affects the sensitivity of the chart to detect a real change. In particular, small values of M are more effective in promptly detecting smaller process shifts. Thus, for monitoring Series A, we decided to use $M = 10$. Observe that raising the prescribed value of p_0 from 0.1 to 0.2 leads to a greater percentage of false alarms but also to a greater power detection of the control chart during out-of-control conditions (results for $p_0 = 0.2$ are also shown in Table 4). Hence, if the practitioner considers acceptable a larger value of p_0 a better out-of-control performance can be achieved.

Table 5: Performance of the stochastic approximation algorithm.

		Sum of Squares			Sum of Squares	
	Total	Estimation	Stochastic	Total	Estimation	Stochastic
n	Mean	Errors	Approximation	Mean	Errors	Approximation
		Model M1			Model M2	
50	24.97	36556.63	73.12	27.06	101531.50	63.89
100	17.71	2641.26	1.45	17.67	5366.76	1.73
300	14.70	95.12	0.60	14.44	299.20	0.59
		Model M3			Model M4	
50	30.50	102578.50	51.80	28.88	153366.42	87.98
100	18.82	862.10	1.92	18.70	2140.08	1.51
300	14.84	80.26	0.61	14.86	86.84	0.56

Note. Each entry is based on 200 time series of length n (used to fit the model) and 20 runs of the Polyak-Ruppert algorithm starting from a random point (used to estimate the within sum of squares).

Phase II. Figure 4 shows the $GLRW_{OS}$ statistic applied to the residuals numbered from 151 to 197. To simulate an out-of-control situation, a shift of magnitude equal to one standard deviation has been added to the original observations from time $\tau = 191$ to 197. Observe that the chart triggers an alarm at time 192, that is one observation later than the true change time. The m.l.e. of τ, δ and ν are equal to 191, 1.334 and 1, respectively. Hence, the $GLRW_{OS}$ correctly identifies the time of change and suggests a possible shift in the process mean.

6 Monte Carlo experiments

We have performed a Monte Carlo experiment to investigate the Polyack-Ruppert algorithm reliability and to which extent the bootstrap control limits are able to guarantee the prescribed frequency of false alarms.

Regarding the use of stochastic approximation we have generated 200 time series of length 50, 100 and 300, for each of the model M1-M4 introduced in Section 3. Then, for each time series, we have indipendently computed 20 solutions of equation (18) for a $GLRW_{OS}$ control chart based on $M = 20$. All the solutions have been computed setting $N_0 = 100$, $p_0 = 0.1$, $A = 20$, $\alpha = 0.6$, $s_0 = 100$, $s = 10000$ and randomly choosing the starting point of the iterative algorithm, h_1, between 10 and 50. For each of the considered cases, Table 5 shows the average of the 200×20 h estimates and, using the 200 time series as grouping factor, the standard decomposition of the sum of squares (SS) into a between and within sum of squares. Here, in particular, the between and within SS measure the variability due to the estimation errors and to the stochastic approximation, respectively. Overall, the variability due to stochastic approximation is much smaller than that due to the estimation errors (the within variability is always less than the 1% of the total variability). The within-group standard errors range from 0.012, for $n = 300$, to a

Table 6: Probabilities of false signal within $t = 100$ of GLR charts ($M = 20$) designed by bootstrap using $N_0 = 100$ and $p_0 = 0.1$.

	$GLRW_{OS}$ charts		$GLRW_{OA}$ charts		$GLRW_{C}$ charts	
n	FS_{ave}	$FS_{\leq 0.2}$	FS_{ave}	$FS_{\leq 0.2}$	FS_{ave}	$FS_{\leq 0.2}$
			Model M1			
50	0.099	0.832	0.101	0.830	0.095	0.843
100	0.102	0.851	0.103	0.847	0.099	0.854
300	0.097	0.935	0.097	0.927	0.095	0.935
			Model M2			
50	0.097	0.872	0.096	0.871	0.095	0.863
100	0.101	0.873	0.098	0.873	0.103	0.870
300	0.104	0.887	0.104	0.883	0.102	0.892
			Model M3			
50	0.106	0.832	0.108	0.841	0.105	0.846
100	0.098	0.849	0.099	0.851	0.096	0.864
300	0.097	0.897	0.099	0.898	0.095	0.919
			Model M4			
50	0.154	0.746	0.148	0.755	0.150	0.752
100	0.110	0.822	0.110	0.821	0.108	0.830
300	0.102	0.900	0.102	0.897	0.100	0.908

Note. FS_{ave} denotes the marginal probability, $FS_{\leq 0.2}$ the frequency with which the conditional probabilities $P(rl \leq 100|\hat{\beta}_n)$ are less or equal to 0.2. Each entry has been estimated by simulation using 2000 time series of length n (used to fit the model and to estimate the control limits) and 1000 continuation for each time series (used to estimate the false signal probabilities). Maximum standard errors goes from 0.0055 when $n = 50$ to 0.0018 when $n = 300$ for FS_{ave} and from 0.0097 to 0.0075 for $FS_{\leq 0.2}$.

maximum of 0.152 when $n = 50$. This guarantees, as predicted by the theory, that different indipendent runs of the stochastic approximation algorithm converge to the same solution without regards to the starting point. In addition, other experiments, conducted with A in the range from 10 to 50 and α in the range from 0.55 to 0.70, have confirmed that the choice of these two parameters are largely not influential on the final outcome at least when the number of iteration s is greater than 5000.

To understand the performance of the bootstrap control limits, from each of the four models $M1$-$M4$ introduced in Section 3, 2000 Phase I time series of length $n = 50, 100$, and 300 have been simulated in order to identify the generating model and estimate the control limits. In particular, for $M = 20$, the omnibus and the combined GLR charts have been designed solving for h the quantile criterion (18) with $N_0 = 100$ and $p_0 = 0.1$. For the stochastic approximation algorithm, we have used the previously given setting but always starting from 20. For each of the original trajectories generated during Phase I, 1000 pseudo-random continuations have been generated in the in-control and out-of-control scenarios. In particular, we have assumed that under the synchronous model (3) a jump occurs on the mean/variance

Table 7: Probabilities of false signal within $t = 100$ of GLR charts ($M = 20$) designed using $N_0 = 100$ and $p_0 = 0.1$ for model M4 either assuming the model order known or without taking into account the impact of the estimation errors (naive design).

	$GLRW_{OS}$ charts		$GLRW_{OA}$ charts		$GLRW_{C}$ charts	
n	FS_{ave}	$FS_{\leq 0.2}$	FS_{ave}	$FS_{\leq 0.2}$	FS_{ave}	$FS_{\leq 0.2}$
Model M4 - Bootstrap limits assuming the order known						
50	0.107	0.806	0.104	0.808	0.107	0.804
100	0.106	0.827	0.106	0.831	0.105	0.837
300	0.100	0.908	0.100	0.902	0.098	0.921
Model M4 - Naive control limits						
50	0.504	0.270	0.512	0.267	0.508	0.261
100	0.294	0.463	0.301	0.454	0.294	0.462
300	0.153	0.755	0.155	0.741	0.150	0.764

Note. FS_{ave} denotes the marginal probability, $FS_{\leq 0.2}$ the frequency with which the conditional probabilities $P(rl \leq 100|\hat{\beta}_n)$ are less or equal to 0.2. Each entry has been estimated as in table 6. Maximum standard errors goes from 0.0075 when $n = 50$ to 0.0023 when $n = 300$ for FS_{ave} and from 0.0112 to 0.0098 for $FS_{\leq 0.2}$.

a $t = 101$ whereas, under the asynchronous model (7), the mean (variance) process shifts at $t = 101$ and the variance (mean) at $t = 106$. Then, we have recorded the run-lengths defined by the stopping rules (6), (12) and (15).

Estimates of both the marginal false alarm probability and the percentage of times that conditional probability is smaller than 0.2, are listed in Table 6. Results show that bootstrap control limits seem to achieve the prescribed frequency of false alarms for each combination of ARMA models and Phase I sample sizes, with the only exception of model $M4$ when $n = 50$. In this case, inaccuracies are due to the difficulty to correctly identify the structure of model M4 when the length of the time series is small. Indeed, when $n = 50$, the Hannan-Rissanen procedure correctly identifies the true order in less than 50% of the cases and, on the contrary, identifies the model as an $AR(2)$ in nearly 30% of the situations. Table 7, showing results analogous to Table 6 in the case of an *a priori* known model structure, clearly confirms that the problem with model M4, when $n = 50$, is due to the high fraction of wrong model order choices. Indeed, computing the control limits using the bootstrap and the stochastic approximation procedure, but assuming that the model order is known, seem to overcome the problem. It is also interesting to note that the problem seems more inherent to the model $M4$ structure than to the adopted identification procedure. Indeed, even when the model is wrongly identified, the Box-Ljung test computed for the first five or ten residual autocorrelations, is not significant at the 5% significant level in more than 99% of the time. Hence, in the case of model M4 and using $n = 50$, the correct identification of model structure can be difficult also for a time-series expert; in addition, using selection criteria different from BIC, such as AIC or AICC, only partially ameliorates the problem due to the biased selection. Further, these alternative criteria introduce in the parameter estimation

Table 8: Probabilities of detection within 20 timesteps of a change point occuring at $t = 100$. Model M1.

	$\tau_1 = 101$	$\tau_1 = 101$	$\tau_1 = \infty$	$\tau_1 = \infty$	$\tau_1 = 101$	$\tau_1 = 101$	$\tau_1 = 106$
	$\tau_2 = \infty$	$\tau_2 = \infty$	$\tau_2 = 101$	$\tau_2 = 101$	$\tau_2 = 101$	$\tau_2 = 106$	$\tau_2 = 101$
	$\delta = 1$	$\delta = 2$	$\delta = -$	$\delta = -$	$\delta = 0.5$	$\delta = 0.5$	$\delta = 0.5$
n	$\nu = -$	$\nu = -$	$\nu = 1.5$	$\nu = 2$	$\nu = 1.25$	$\nu = 1.25$	$\nu = 1.25$
				$GLRW_{OS}$ charts			
50	0.246	0.931	0.267	0.762	0.149	0.118	0.130
100	0.402	0.991	0.407	0.885	0.240	0.194	0.210
300	0.530	0.999	0.512	0.934	0.320	0.263	0.282
				$GLRW_{OA}$ charts			
50	0.227	0.918	0.291	0.784	0.154	0.123	0.136
100	0.372	0.988	0.446	0.904	0.251	0.203	0.223
300	0.501	0.998	0.563	0.950	0.340	0.278	0.304
				$GLRW_{C}$ charts			
50	0.218	0.919	0.307	0.805	0.147	0.114	0.133
100	0.358	0.987	0.464	0.915	0.241	0.191	0.219
300	0.481	0.998	0.574	0.954	0.323	0.260	0.294

Note. Each entry has been estimated by simulation as in Table 6. Maximum standard errors goes from 0.0084 when $n = 50$ to 0.0025 when $n = 300$.

process an extra-variability so large to severely reduce the GLR power detection for model M4 and for other ARMA models and, consequently, the wide applicability of the proposed procedure.

Overall, our results show that the bootstrap procedure can be successfully implemented for most of the investigated time series model even using a Phase I sample size of $n = 50$. Even the relatively worse performance of model $M4$ for the smaller sample size is still much better than that obtained using the *naive* control limits, i.e., ignoring the impact of model uncertainty (see Table 7). However, in order to improve the initial performance, we emphasize the importance of updating the control limits, when at least 100 observations become available. Note that, this can be easily done since, given M, N_0 and p_0, the suggested procedure leads to a completely automatic updating.

The need to update the decision interval arises also by the analysis of the out-of-control performance of the GLR charts based on the bootstrap control limits. Table 8, in particular, shows the probabilities to trigger an alarm within 20 timesteps when synchronus or asynchronous change points occur in the mean and/or variance of model $M1$ (similar results have been obtained for the other models). As expected, the GLR power detection strictly depends on the number of observations gathered during the Phase I. Indeed, since the bootstrap decision intervals take into account the estimation errors, and these are smaller when n is larger, the probability to promptly give a true signal increases with the Phase I sample size. Regarding to comparisons between the different version of the window-limited GLR charts, observe that (i) when only the process mean shifts, the omnibus GLR chart based on the synchronous change point model, i.e. $GLRW_{OS}$, slightly outperforms both

the asynchronous and combined GLR charts; (ii) when only the process variance shifts, the combined GLR chart, i.e. $GLRW_C$, shows the better power detection; (iii) when a change affects both the mean and the variance process the GLR chart based on the asynchronous change point model, i.e. $GLRW_{OA}$, slightly outperforms the other two control schemes. However, it should be noted that these differences between the charts are really small and perhaps unimportant in practical situations.

7 Software

We have implemented a library of functions in the R language (R Development Core Team, 2006), with the most time consuming part written in C. Our functions use the Chandresekar recursions to compute the standardized innovations and the likelihood function (Pearlman, 1980; Melard, 1984) and the *newuoa* algorithm, introduced by Powell (2004), to derive the maximum likelihood estimates. Model selection, required at the second stage of the Hannan and Rissanen algorithm, has been implemented using the QR decomposition and moving variables in and out of the model according to the algorithm described by Clarke (1981).

Observe that the $GLRW$ statistics are functions of the following basic quantities:

$$w_t(i) = \sum_{h=t-j}^{t} a_h^2, \;\; c_t(i,j) = \sum_{h=t-j}^{t} a_h \rho(h, t-i) \text{ and } v_t(i,j) = \sum_{h=t-j}^{t} \rho^2(h, t-i) \quad (25)$$

for $i = 0, \ldots, M-1$ and $j = 0, \ldots, i$. For example, regarding to the $GLRW_{OS}$ statistics, we have that

$$\hat{\delta}_t(\tau) = \frac{c_t(t-\tau, t-\tau)}{v_t(t-\tau, t-\tau)} \;\text{ and }\; \hat{\nu}_t(\tau) = \max(1, s_t^2(\tau, \hat{\delta}_t(\tau)))$$

where $s_t^2(\tau, \hat{\delta}_t(\tau)) = (w_t(t-\tau) - \hat{\delta}_t^2(\tau) v_t(t-\tau, t-\tau))/(t-\tau+1)$. Hence,

$$\sup_{\delta, \nu^2 \geq 1} l_1(\tau, \delta, \nu^2) = l_1(\tau, \hat{\delta}_t(\tau), \hat{\nu}_t(\tau)) = w_t(t-\tau) - (t-\tau+1)\{s_t^2(\tau, \hat{\delta}_t(\tau))/\hat{\nu}_t^2(\tau) + \log(\hat{\nu}_t^2(\tau))\}$$

Note that the previous quantities can be efficiently computed using the following recursive relations

$$\begin{aligned} &w_t(i) = w_t(i-1) + a_t^2 \\ &c_t(i,j) = c_t(i-1, j-1) + a_t \rho(t, t-i) \qquad i = 1, \ldots, M-1; j = 1, \ldots, i \\ &v_t(i,j) = v_t(i-1, j-1) + \rho^2(t, t-i) \end{aligned}$$

with $w_t(0) = a_t^2$, $c_t(i,0) = a_t \rho(t, t-i)$, $v_t(i,0) = \rho^2(t, t-i)$ for $i = 0, \ldots, M$. The use of these recursions is particularly useful for the asynchronous GLR implementation. When more than one window-length value is considered, the previous recursions must be runned using the maximum M. Then, all the GLRW statistics can be readily computed for different values of M.

In order to give an example of the computing time, the 10100 stochastic approximation iterations, used in the example of Section 5 to compute the six decision

intervals, for $M = 5, 10, 15, 20, 25$ and 30, took a total of 19 seconds using a Pentium 4 (3.00GHZ) PC running the GNU/Linux (Ubuntu 6.10) operating system. Observe that, in this case, 10101 ARMA models have been fitted to time series of 150 observations using automatic order selection and exact maximum likelihood estimation.

8 Conclusions

In this paper we have proposed a procedure to design a GLR chart with a prescribed rate of false alarms. The methodology can be summarized as follows:

a) fit an ARMA model to the data using the Hannan and Rissanen (1982) automatic identification procedure;
b) check model assumptions; in particular, i) check the Gaussian assumption using a suitable test and normal probability plot; ii) check for no autocorrelation in the residuals plotting the sample autocorrelation and, if necessary, using some test like the Ljung-Box portmanteau test (Wei, 2006); iii) check for the presence of outliers and/or level changes;
c) choose N_0 and p_0, or a suitable large value of the in-control ARL; then compute, for different values of the window size M, the control limit h using stochastic approximation;
d) choose a suitable value of M on the basis of some estimate of the out-of-control performance.

We suggest starting the designing procedure when at least 50 Phase I observations are available and to regularly update the control limits as more data are gathered. If needed, the design procedure can be implemented in a fully automatic way. Indeed, step b) can be automated using a predefined battery of diagnostic tests and the choice of M in step d) can be based on a prefixed criteria (e.g. M might be chosen so as to maximize the probability of detecting a given shift within r timesteps). However, we believe that, if it is possible, users should have a minimal interaction with the designing procedure in order to make more balanced choices and also to achieve a better understanding of the underlying process. In any way, if suitable software is available, the interactive design for a particular process requires only a few minutes.

Note that, it is also necessary to choose between implementing a synchronous, an asynchronous or a combined GLR. However, our results show little practical differences between these three types of GLR charts in terms of in and out-of-control performance. Since their statistical properties are basically equivalent, we emphasize that this choice should be mainly based on the practitioner's preference and/or experience.

Finally, note that the proposed approach can be also used for the design of other control charts. For example, we have successfully applied a variant of the presented procedure to the design of residual EWMA and CUSUM monitoring schemes.

References

Adams, B. and Tzeng, I. (1998). Robustness of forecast-based monitoring schemes. *Journal of Quality Technology*, 30:328–339.

Alwan, L. and Roberts, H. (1988). Time-series modelling for statistical process control. *Journal of Business and Economic Statistics*, 6:87–95.

Apley, D. (2002). Time series control charts in the presence of model uncertainty. *Journal of Manifacturing Science and Engineering*, 124:891–898.

Apley, D. and Lee, H. (2003). Design of exponentially weighted moving average control charts for autocorrelated processes with model uncertainty. *Technometrics*, 45:187–198.

Apley, D. and Shi, J. (1999). The GLRT for statistical process control of autocorrelated processes. *IIE Transactions*, 31:1123–1134.

Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes, Theory and Applications*. Prentice Hall, Englewood Cliffs, N.J.

Box, G. E. P., Jenkins, G., and G.C., R. (1994). *Time Series Analysis Forecasting and Control*. Prentice Hall, New Jersey, 3nd edition.

Boyles, R. (2000). Phase I analysis for autocorrelated processes. *Journal of Quality Technology*, 32:395–409.

Brockwell, P. J. and Davies, R. A. (1996). *Time Series: Theory and Methods*. Springer, New York, 2nd edition.

Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17:52–72.

Chang, J. and Fricker, R. (1999). Detecting when a monotonically increasing mean has crossed a threshold. *Journal of Quality Technology*, 31:217–234.

Clarke, M. (1981). Algorithm AS 163: A Givens algorithm for moving from one linear model to another without going back to the data. *Applied Statistics*, 30:198–203.

Gan, F. (1993). An optimal design of EWMA control charts based on median run length. *Journal of Statistical Computation and Simulation*, 45:169–184.

Gan, F. (1994). An optimal design of cumulative sum control charts based on median run length. *Communication in Statistics-Simulation and Computation*, 23:485–503.

Hannan, E. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69:81–94.

Harris, T. and Ross, W. (1991). Statistical process control procedures for correlated observations. *Canadian Journal of Chemical Engineering*, 69:48–57.

Jensen, W., Jones-Farmer, L., Champ, C., and Woodall, W. (2006). Effects of parameter estimation on control charts properties: a literature review. *Journal of Quality Technology*, 38:349–364.

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

Jones, L., Champ, C., and Rigdon, S. (2001). The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*, 42:156–167.

Kavalieris, L. (1991). A note on estimating autoregressive-moving average order. *Biometrika*, 78:920–922.

Kramer, H. and Schmid, W. (2000). The influence of parameter estimation on the ARL of Shewhart type charts for time series. *Statistical Papers*, 41:173–196.

Kushner, H. and Yin, G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York.

Lai, T. (1995). Sequential change-point detection in quality control and dynamical systems. *Journal of Royal Statistical Society, B*, 57:613–658.

Lai, T. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, 11:303–408.

Lee, M. and Khoo, M. (2006). Optimal statistical design of a multivariate EWMA chart based on ARL and MRL. *Communication in Statistics-Simulation and Computation*, 35:831–437.

Lu, C. and Reynolds, M. (1999a). Control charts for monitoring the mean and variance of autocorrelated processes. *Journal of Quality Technology*, 31:259–274.

Lu, C. and Reynolds, M. (1999b). EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology*, 31:166–188.

Melard, G. (1984). Algorithm AS 197: a fast algorithm for the exact likelihood of autoregressive-moving average models. *Applied Statistics*, 33:104–114.

Montgomery, D. and Mastrangelo, C. (1991). Some statistical process control methods for autocorrelated data (with discussion). *Journal of Quality Technology*, 23:179–274.

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

Pearlman, J. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika*, 67:232–233.

Polyak, B. and Juditsky, A. (1992). Accelaration of stochastic approximation by averaging. *SIAM Journal of Control Optmization*, 30:838–855.

Powell, M. (2004). The NEWUOA software for uncostrained optimization without derivatives. Technical Report DAMTP 2004/NA08, CMS Building, Wilberforce Road, Cambridge CB3 0WA, UK.

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

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Runger, G. (2002). Assignable causes and autocorrelation: Control charts for observations or residuals? *Journal of Quality Technology*, 34:165–170.

Runger, G. and Testik, M. (2003). Control charts for monitoring fault signatures: Cuscore versus GLR. *Quality and Reliability Engineering International*, 19:387–396.

Runger, G. and Willemain, T. (1995). Model-based and model-free control of autocorrelated processes. *Journal of Quality Technology*, 27:283–292.

Runger, G., Willemain, T., and Prabhu, S. (1995). Average run lengths for CUSUM control charts applied to residuals. *Communications in Statistics-Theory and Methods*, 24:273–282.

Ruppert, D. (1991). Stochastic approximation. In Ghosh, B. and Sen, P., editors, *Handbook of Sequential Analysis*, pages 503–529. Marcel Dekker, New York.

Schimd, W. and Schöne, A. (1997). Some properties of the EWMA control chart in the presence of autocorrelation. *The Annals of Statistics*, 25:1277–1283.

Shu, L., Apley, D., and Tsung, F. (2002). Autocorrelated process monitoring using triggered CUSCORE charts. *Quality and Reliability Engineering International*, 18:411–421.

Siegmund, D. and Venkatraman, E. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Annals of Statistics*, 23:255–271.

Superville, C. and Adams, B. (1994). An evaluation of forecast-based quality control schemes. *Communication in Statistics: Simulation and Computation*, 23:645–661.

Testik, M. (2005). Model inadequacy and residual control charts for autocorrelated processes. *Quality and Reliability Engineering International*, 21:115–130.

Vasilopoulos, A. and Stamboulis, A. (1978). Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology*, 10:20–30.

Wardell, D., Moskowitz, H., and Plante, R. (1994). Run length distributions of special cause control charts for correlated processes. *Technometrics*, 36:3–27.

Wei, W. (2006). *Time Series analysis univariate and multivariate methods*. Addison-Wesley Publishing Company, Redwood City, California, 2nd edition edition.

Willsky, A. and Jones, H. (1976). A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. *IEEE Transactions of Automatic Control*, 21:108–112.

Yashchin, E. (1993). Performance of CUSUM control schemes for serially correlated observations. *Technometrics*, 35:35–52.

Acknowledgements

This research was partially funded by Italian MIUR-Cofin 2006 grants.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy