2003

Applications of recursive estimation methods in statistical process control

Shelton Peiris

Aerambamoorthy Thavaneswaran

David E. Allen

R. Mellor

Applications of Recursive Estimation Methods in Statistical Process Control

By

S. Peiris*
Department of Statistics, Pennsylvania State University

A. Thavaneswaran
Department pf Statistics, University of Manitoba

D. Allen
School of Accounting, Finance and Economics,
Edith Cowan University

R. Mellor
School oof Quantitative Methods, University of Western Sydney

*Correspondence author and address:

S. Peiris
Department of Statistics
Pennsylvania State University
University Park PA 16802
U.S.A.

Abstract

In recent years there has been a growing interest in recursive estimation techniques as applied to statistical process control (SPC). In cases where prior information about the processes are available, it is shown that procedures based on the "optimal" smoothing can be superior to the classical procedures like Shewhart's CUSUM control charts (see, for instance, Thavaneswaran, McPherson and Abraham (1998)). This paper reviews the recursive algorithms based on EWMA (exponentially weighted moving average), DLM (dynamic linear modeling), KF (Kalman filtering) and OS (optimal smoothing) in statistical process control with correlated data. We also discuss various relationships among the asymptotic mean square errors (MSE) of these procedures in SPC.

Key words: Process control; time series, Kalman filter; estimation; correlated data; control charts; optimal smoothing; non-Gaussian; EWMA; recursive

1. Introduction

The theory and applications of control charts are very useful for improving quality and productivity in the resolution of many industrial problems. This theory has been developed on the basis of a fundamental assumption that the observations in the process are independent (see Hunter (1986)). In this situation it is reasonable to model the observations using a simple model such that

$$X_t = \mu + \varepsilon_t, \tag{1}$$

where X_t denotes the tth $(t = 1, \cdots, n)$ observation, μ is the process mean and ε_t's are independent and identically distributed (iid) random variables having zero mean and constant variance σ^2. In standard applications of (1), there are three popular types of control charts, namely, the Shewhart, the cumulative sum (CUSUM) and the EWMA charts. The latter, EWMA chart is constructed using the fact that the one-step-ahead forecast function based on (1) is given by the exponential smoothing scheme satisfying

$$Z_t = \hat{X}_{t,1} = \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j}, \tag{2}$$

where $0 < \lambda < 1$. Equation (2) is equivalent to the recursion

$$Z_t = \lambda X_t + (1-\lambda)Z_{t-1}, \tag{3}$$

where Z_0 is a suitable starting-up value.

This EWMA suggested by Roberts (1959) is a very popular approach in SPC. If the observations are uncorrelated as given in (1) Hunter (1986), Montgomery and Mastrangelo (1991) have shown that the control limits for an EWMA control chart (under steady state conditions) are given by

$$\text{(Upper Control Limit) UCL} = \overline{X} + K\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}} \quad \text{and}$$

$$\text{(Lower Control Limit) LCL} = \overline{X} - K\sigma\sqrt{\frac{\lambda}{n(2-\lambda)}},$$

where $\overline{X}$ is the sample mean. Note that typically $K = 2$ for 'warning limits' and $K = 3$ for 'out of control' situations. However, it is known that, in practice, the assumption of independence is often violated. Detection of autocorrelation can be accomplished through diagnostic plots or through formal statistical tests. A simple plot of the residuals from a model can be helpful. For example, if the residuals are plotted against time, rapid changes in sign may indicate the

presence of a negative autocorrelation. On the other hand, an unusually large number of residuals clustered together on the same side of the mean indicate that the errors are governed by positive autocorrelation. When the assumption of independence is violated there is a serious impact on the control charts due to the presence of this autocorrelation. If the effect of the existence of positively (or negatively) correlated errors is ignored, then an estimate of σ^2 is clearly a substantial underestimate (overestimate) of the true variance. Generally, this results in an increase in the frequency of false alarms. In other words, the Average Run Length (ARL) is much shorter than it would be for a process with uncorrelated observations. Thus the true state of control of the process often cannot be determined from standard control charts.

With that view in mind, Section 2 reviews the theory of autoregressive integrated moving average processes of order (p,d,q) (ARIMA (p,d,q)), EWMA and DLM procedures in brief in order to accommodate the autocorrelation structure of data. Section 3 considers the recursive estimation and predication algorithms based on Bayes and KF techniques for correlated observations. Section 4 investigates the effect of the MSE of these procedures and compares them using numerical examples.

2.   Models for Autocorrelated Data

In literature there are three main methods available for constructing control charts for correlated data. The first method is based on the fitting of an autoregressive integrated moving average (ARIMA) model to the original data and then constructing control charts for the residuals. The second method uses a control chart based upon the EWMA statistic and the third method is based on the dynamic structure of the underlying process.

2.1   ARIMA (p,d,q) Models

The ARIMA (p,d,q) is given by

$$\Phi(B)(I-B)^d X_t = \Theta(B)\varepsilon_t, \tag{4}$$

where $\Phi(B) = I - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, and $\Theta(B) = I - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ are stationary autoregressive (AR) and invertible moving average (MA) polynomials in the backshift

operator $B\left(B^j X_t = X_{t-j}\right)$ of orders p and q respectively and ε_t's are identically distributed independent random variables with mean 0 and variance σ^2.

Let $\hat{X}_t$ be the predicated values from a fitted ARIMA model in (4). Then the one-step-ahead prediction errors (or residuals) are given by

$$\varepsilon_t = X_t - \hat{X}_t; t = 1,2,\cdots.$$

Recall that if the fitted model is adequate, ε_t's are identically distributed independent random variables. Hence for SPC with correlated data, one can apply control charts to the stream of estimated residuals from a fitted ARIMA (p,d,q) model (see, for example, Alwan and Roberts (1988) and Montogomery and Mastrangelo (1991)). The ARIMA process $\{X_t\}$ in (4) is said to be 'out of control" if there is a significant shift in the mean level. This shift can be detected from the control chart applied to the residuals since the model misspecification is assumed to be transferred to the residuals.

2.2   EWMA Charts for Autocorrelated Data

The EWMA scheme in (1.3) can also be used for autocorrelated data generated by some specific models. As an illustration, suppose that a set of nonstationary (homogeneous) time series data can be modeled by an ARIMA(0,1,1,) or IMA(1,1) process given by

$$U_t = \varepsilon_t - \theta\varepsilon_{t-1}, \tag{5}$$

where $U_t = X_t - X_{t-1}$. Let $Z_t = \hat{X}_t(1)$ be the one-step-ahead forecast (in the minimum mean square sense) for time $t+1$ made at time t. The corresponding optimal forecast function is given by the conditional mean

$$Z_t = E\left(X_{t+1}|X_t, X_{t-1},\cdots\right) = X_t - \theta\varepsilon_t. \tag{6}$$

This is equivalent to the equation (3) with $\lambda = 1 - \theta$ since $\varepsilon_t = X_t - Z_{t-1}$. In this case, the one step ahead forecast errors ε_t's are independent with men zero and standard deviation σ if the fitted IMA(1,1) model is adequate. Thus we could set up control charts for the one step ahead forecast errors.

Montgomery and Mastrangelo (1991) argued that an EWMA scheme with a suitably chosen λ will give an excellent one step ahead forecast even if

(a)   the observations from a process are positively autocorrelated and

(b)   the process mean does not drift too quickly.

Recently there has been a growing interest in the theory and application of recursive and filtering techniques in SPC (see, for instance, Thavaneswaran et al. (1998)). Section 2.3 considers a first order dynamic linear model in which the mean is a slowly changing random walk and show that this could be well represented by an EWMA scheme with suitably chosen parameters.

2.3   First Order Dynamic Linear Models

Suppose that the observation series is generated by
$$X_t = \mu_t + v_t, \tag{7}$$
where $v_t \sim N\left(0, U_t^2\right)$ and μ_t is the mean level at time t. Let the time evolution of the mean level follow a simple random walk
$$\mu_t = \mu_{t-1} + \omega_t, \tag{8}$$
where $\omega_t \sim N\left(0, V_t^2\right)$. Assume that $\{v_t\}$ and $\{\omega_t\}$ are independent over the time and are mutually independent processes for all t. An important special case of (7) and (8) can be obtained by setting the variances to be constants i.e. $U_t^2 = U^2$ and $V_t^2 = V^2$ for all t. The relative variation of the state in (8) to the observation equation (7) in this latter constant variance case is given by $r = \dfrac{V^2}{U^2}$ and is known as the signal to noise ratio. In this constant variance case (7) and (8) reduce to
$$X_t - X_{t-1} = \omega_t + v_t - v_{t-1}. \tag{9}$$
 Clearly, (9) can be written as an IMA(1,1) model for suitably chosen parameters θ and σ^2 satisfying
$$X_t - X_{t-1} = \varsigma_t - \theta\varsigma_{t-1}, \tag{10}$$
where θ is the solution of $U^2\theta^2 - \theta\left(2U^2 + V^2\right) + U^2 = 0$ such that $|\theta| < 1$ and ς_t is a suitably chosen NID (identically distributed, independent normal noise process with variance U^2/θ . The prediction from a constant variance dynamic linear model satisfying (7) and (8) can be obtained by the EWMA algorithm given in Section 2.2. In Section 3, we discuss two alternative recursive estimation and prediction methods based on Bayesian and Kalman filtering algorithms.

3. Recursive Estimation via Bayesian and Kalman Filtering Methods

We first consider the Bayesian method using the first order DLM discussed in Section 2.3.

3.1 The Bayesian Method

Let $\Im^t$ be the σ algebra generated by $X_t, X_{t-1}, \cdots, X_1$ and let $\Im^0$ be the available initial information. Suppose that the initial distribution, $\left(\mu_0 \middle| \Im^0\right) \sim N\left(m_0, \alpha_0^2\right)$ and the posterior distribution, $\mu_{t-1} \middle| \Im^{t-1} \sim N\left(m_{t-1}, \alpha_{t-1}^2\right)$ are known. Now we have the following theorem:

Theorem:

(a) The prior distribution, $\left(\mu_t \middle| \Im^{t-1}\right) \sim N\left(m_{t-1}, \alpha_{t-1}^2 + U^2\right)$,

(b) The forecast distribution, $\left(X_t \middle| \Im^{t-1}\right) \sim N\left(m_{t-1}, \alpha_{t-1}^2 + U^2 + V^2\right)$,

(c) The posterior distribution, $\left(\mu_t \middle| \Im^t\right) \sim N\left(m_t, \alpha_t^2\right)$,

where the one-step-ahead forecast function satisfies the recursion

$$m_t = m_{t-1} + \frac{\alpha_{t-1}^2 + U^2}{\alpha_{t-1}^2 + U^2 + V^2}\left(X_t - m_{t-1}\right).$$

See West and Harrison (1989) for details. An alternative proof of similar results without the normality assumptions can be found in Thavaneswaran et. al. (1998). Section 3.2 considers the Kalman filtering approach to an ARIMA (0,1,1) model for the sake of a comparison of recursive algorithms.

3.2 Recursive Estimation via Kalman Filtering

In forecasting, the recursive nature of the basic equations cannot be neglected. One popular method of incorporating this recursive nature is known as Kalman filtering (KF), as originally suggested by Kalman (1960). In recent years the state-space representations and KF algorithms for certain processes have played an important role in estimation and prediction problems. The development of this theory consists of two equations, known as 'the observation equation' and 'the state equation' as given below:

Suppose that the observation equation for the scalar case is

$$Y_t = G_t X_t + v_t, \tag{11}$$

where G_t and X_t are suitably chosen vectors and $v_t \sim N(0, U^2)$. The time evolution of the state vector X_t follows the state equation

$$X_{t+1} = F_t X_t + \omega_t, \tag{12}$$

where F_t is a suitable matrix and $\omega_t \sim N(0, V^2)$. Consider an ARIMA(0,1,1) model given in equation (10) for illustration. We write (10) using the above state-space form based on the following vectors and matrices:

$$F_t = (-\theta, ,1, ,1) = F,$$

$$X_t = (X_{t-1,} X_{t,} Y_{t-1})'.$$

Applying the standard Kalman prediction algorithm (see, for instance, Brockwell and Davis 1996, p265) one has the following recursion for the one-step-ahead forecast function, Z_t:

$$Z_t = a X_t + (1 - a) Z_{t-1}, \tag{13}$$

where $a = \dfrac{\Omega}{\Omega + U^2}$ and $\Omega = \left\{ V^2 + \sqrt{V^4 + 4U^2 V^2} \right\}\!/2$

Section 4 considers a numerical illustration and investigates the relative asymptotic efficiency of the methods described in Sections 2 and 3.

4.   A Numerical Illustration

In the equilibrium situation for large t, it is assumed that the gain in information from a new observation is exactly balanced by the loss of information as we pass by from one state to the next. We summarize the following asymptotic MSE's (under steady state conditions) in each case as given below:

<div align="center">Asymptotic MSE Summary</div>

Method	Asymptotic MSE	Note
EWMA	$\lambda \sigma^2 / (2 - \lambda)$	$\lambda = 1 - \theta$
DLM	$(A+1)U^2 + V^2$	$A = \left\{ V\left(\sqrt{V^2 + 4U^2} - V \right) \right\}\!/2$
Bayes	AU^2	
Kalman	$\Omega + U^2$	$\Omega = \left\{ V^2 + \sqrt{V^4 + 4U^2 V^2} \right\}\!/2$

Next we consider a simulation study and compare the relevant asymptotic mean square errors (MSE) in each case. Consider a simulation study based on the models in (7) and (8). We set the following parameter values:

$$V^2 = 1, U^2 = 4, \mu_0 = 0 \text{ and the distributions,}$$

$$v_t \sim N(0,4), \omega_t \sim N(0,1).$$

First, simulate 100 values from $\mu_t = \mu_{t-1} + \omega_t$ with $\mu_0 = 0$. Then 100 values of the observation series is generated from $X_t = \mu_t + v_t$. The series X_t and the random walk for the mean are given in Fig. 1. The autocorrelation function of X_t and $X_t - X_{t-1}$ are also shown in Fig 1. Since the difference series $X_t - X_{t-1}$ follows an MA(1) process, we fit the model $X_t - X_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1}$ to the differenced data. The corresponding parameter estimates are $\hat{\theta} = 0.61$ and $\hat{\sigma}^2 = 6.56$. The following table summarizes the asymptotic MSE's in each case:

Method	Asymptotic MSE
EWMA	1.59
DLM	11.24
Bayes	6.24
Kalman	6.56

As expected, it is clear that the EWMA provides the smallest MSE among these 4 approaches. The main reason for this is that the data is well approximated by an IMA(1,1) model. However, Bayes method provides the smallest MSE among other three approaches. When $\theta \approx 0$, the EWMA, Bayes and KF approaches give very similar MSE's. However, the use DLM in these cases results in a loss of almost 50% efficiency compared to the other two methods based on the Bayes and Kalman approaches.

Fig 1 : Random Walk (unbroken) and Data (broken)
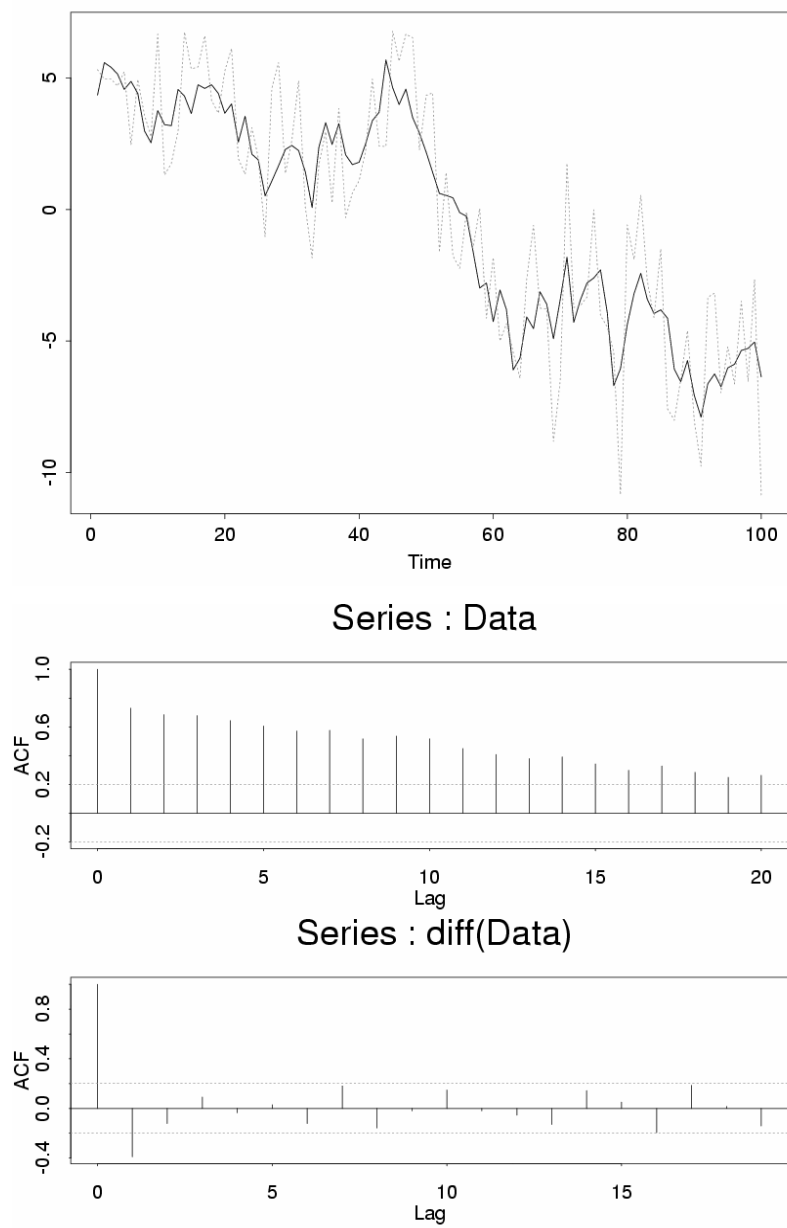
References

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

Abraham, B., Thavaneswaran, A. and Peiris, S., On the predication for a nonlinear time series models using estimating equations. *Lecture Notes Monograph Series*, Vol.32, (1996)pp 259-267. (I. Basawa and R.L. Taylor Eds).

Abranham, B. and J. Ledolter , Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. *International Statistical Review*, 54(1), (1986) pp.5-66.

Brockwell, P.J. and Davis, R.A.,*An Introduction to Time Series and Forecasting, Second Edition*, Springer Verlag, NewYork (1996).

Hunter, J.S., The Exponentially Weighted Moving Average. *Journal of Quality Technology*, Vol.18, No.4 (1986)pp. 203-209.

Kalman, R.E., Anew approach to linear filtering and prediction problems. *J. Basic Engineering* 82, (1960)pp.35-45.

Lowry, C.A. and Montgomery, C.D., A review of multivariate control charts. *IEE Transactions*, 27, (1995) pp.800-810.

Montgomery, D.C. and Mastgomery, C.D. , Some statistical process control methods for autocorrelated data. *Journal of Quality Technology,* July 1991, Vol.23, No.3, (1995)pp.179-268.

Peiris, M.S. , Improving the Quality of Forecasting using Generalized ARMA Models: An Application to Statistical Quality Control. *International conference on Statistics in Industry and Business*, January 2-4, Cochin University, India (2003).

Roberts, S.W., Control Chart Tests Based on Geometric Moving Averages. , *Technometrics*,1 (1959)pp.239-251.

Thavaneswaran, A., Macpherson, B.D., Abraham, B., An Application of Filtering to Statistical Process Control. *Proceedings of the International, Conference on Statistics in Industry and Business*, January 1998, Cochin University India (1998).

West, M. and Harrison, J. , *Bayesian forecasting and Dynamic Modeling*, Springer Verlag, New York (1989).