ROBUST DESIGN OF CONTROL CHARTS FOR

AUTOCORRELATED PROCESSES WITH MODEL

UNCERTAINTY

A Dissertation

by

HYUN CHEOL LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Industrial Engineering

ROBUST DESIGN OF CONTROL CHARTS FOR

AUTOCORRELATED PROCESSES WITH MODEL

UNCERTAINTY

A Dissertation

by

HYUN CHEOL LEE

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

_____
Daniel W. Apley
(Co-Chair of Committee)

_____
Yu Ding
(Co-Chair of Committee)

_____
Michael T. Longnecker
(Member)

_____
Sheng-Jen "Tony" Hsieh
(Member)

_____
Mark L. Spearman
(Head of Department)

August 2004

Major Subject: Industrial Engineering

ABSTRACT

Robust Design of Control Charts for Autocorrelated Processes with Model

Uncertainty. (August 2004)

Hyun Cheol Lee, B.S., Korea University, Korea;

M.S., Texas A&M University

Co-Chair of Advisory Committee: Dr. Daniel Apley
Dr. Yu Ding

Statistical process control (SPC) procedures suitable for autocorrelated processes have been extensively investigated in recent years. The most popular method is the residual-based control chart. To implement this method, a time series model, which is usually an autoregressive moving average (ARMA) model, of the process is required. However, the model must be estimated from data in practice and the resulting ARMA modeling errors are unavoidable. Residual-based control charts are known to be sensitive to ARMA modeling errors and often suffer from inflated false alarm rates. As an alternative, control charts can be applied directly to the autocorrelated data with widened control limits. The widened amount is determined by the autocorrelation function of the process. The alternative method, however, also cannot be free from the effects of modeling errors because it relies on an accurate process model to be effective.

To compare robustness to the ARMA modeling errors between the preceding two kinds of methods for control charting autocorrelated data, this dissertation investigates the sensitivity analytically. Then, two robust design procedures for residual-based control charts are developed from the result of the sensitivity analysis. The first approach for robust design uses the worst-case (maximum) variance of a chart statistic to guarantee the initial specification of control charts. The second robust design method uses the expected variance of the chart statistic. The resulting control limits are widened by an amount that depends on the variance of the chart statistic — maximum or expected — as a function of (among other things) the parameter estimation error covariances.

To my parents for their love, sacrifice, support …and everything they gave me

ACKNOWLEDGMENTS

I would like to express my honest and deepest appreciation to my advisor, Dr. Daniel W. Apley, who has encouraged, inspired and motivated me throughout my graduate studies at Texas A&M. Without your generous advice, support and bountiful knowledge, I could have not accomplished this work. I was so fortunate to have had you as my advisor. I will always admire you.

I would also like to thank the members of my committee: Co-chair Dr. Yu Ding, Dr. Michael T. Longnecker and Dr. Sheng-Jen "Tony" Hsieh. They gave me great advice and were a great influence throughout this long effort.

I want to express appreciation to all my friends in our data analysis laboratory for their warm friendship and valuable advice.

I am also grateful to my family. Special thanks to my parents, Dr. Hough Joon Lee and Mrs. Ae Ja Jeon, who have encouraged me every step of my life. Thanks to my brothers, Yong Doo Lee and Sung Han Lee, and my sister, Yong Kyung Lee, for their love and prayers. Thanks also to my precious nephews, Jae Ho Lee, Jee Min Lee and Jae Eun Kim, for their cute relief at times.

Finally, I thank God for giving me this wonderful opportunity.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

A control chart is one of the primary techniques in statistical process control (SPC) procedures. They are widely used to monitor processes and detect shifts in key quality-related variables. Through the effective implementing control charts in industrial processes, the product quality can be improved. Traditional control charts are based on the assumption that process data are independent. Significant advances in measurement and data collection technology — particularly in the area of in-process sensing — have created the potential for much more frequent inspection. As a result, autocorrelated data are now common (Montgomery and Woodall, 1997; Woodall and Montgomery 1999). The run length properties of traditional control charts like cumulative sum (CUSUM) and $\overline{X}$ charts are strongly affected by data autocorrelation, and the in-control average run length (ARL) can be much shorter than intended if the autocorrelation is positive (Johnson and Bagshaw, 1974; Vasilopoulos and Stamboulis, 1978). Consequently, there has been considerable research in recent years on designing control charts suitable for autocorrelated processes (see, e.g., Montgomery and Woodall, 1997, Lu and Reynolds, 1999, and the references therein).

This dissertation follows the style and the format of *Technometrics*.

I.1 Motivation of the Study

There are two primary classes of methods for control charting autocorrelated data. The first class of methods is residual-based control charts (e.g., Alwan and Roberts, 1988; Apley and Shi, 1999; Berthouex, Hunter, and Pallesen, 1978; English, Krishnamurthi, and Sastri, 1991; Lin and Adams, 1996; Lu and Reynolds, 1999; Montgomery and Mastrangelo, 1991; Runger, Willemain, and Prabhu, 1995; Superville and Adams, 1994; Vander Wiel, 1996; Wardell, Moskowitz, and Plante, 1994). One usually assumes the process data x_t (t is a time index) follows an autoregressive moving average (ARMA) model with AR order p and MA order q, denoted ARMA(p,q). Using standard time series notation (see Box, et al., 1994) with the backward shift operator B defined such that $Bx_t = x_{t-1}$, an ARMA model can be written as

$$x_t = \frac{\Theta(B)}{\Phi(B)} a_t, \tag{1.1}$$

where $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_q B^q$, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p$, and a_t is an independently, identically distributed (iid), 0-mean sequence of random shocks with variance σ_a^2.

The basic idea behind residual-based charts is to directly monitor the residuals (the one-step-ahead prediction errors), generated via $e_t = \Theta^{-1}(B)\Phi(B)x_t$. From (1.1), e_t is exactly the iid sequence a_t, after any initial transients have died out. Thus, traditional Shewhart, CUSUM, and exponentially weighted moving average (EWMA) control

charts can be applied to the uncorrelated residuals with well understood in-control run length properties. Then, residual-based control charts detect a mean shift of the original autocorrelated process as recognizing the mean shift "signature" in the residual process (Apley and Shi 1999).

In practice, however, the model parameters must always be estimated from process data. One criticism of residual-based charts is that they lack robustness to ARMA modeling errors (e.g., Adams and Tseng 1998; Apley and Shi 1999; Lu and Reynolds 1999). For example, since the EWMA is a weighted average of the past residuals, residual autocorrelation due to estimation errors can have a substantial effect on EWMA variance and the resulting in-control ARL. If the true and estimated parameters are such that the residual autocorrelation is positive, the in-control ARL will be shorter than intended, and the control chart may be plagued with frequent false alarms. Illustrative examples are included in Chapters II.4 and III.1.

In the second primary class of methods, a traditional control chart is applied directly to x_t, but the control limits are modified (usually widened) to take into account the autocorrelation. Johnson and Bagshaw (1974), Vasilopoulos and Stamboulis (1978), and Zhang (1998) discussed modifying the control limits of CUSUM charts, $\overline{X}$ charts, and EWMA charts, respectively. The extent to which the control limits are widened depends on the autocorrelation function or, equivalently, on the parameters of the ARMA model used to represent the autocorrelation. To be implemented effectively, this approach also relies on an accurate ARMA process model (or, equivalently, the autocorrelation function of x_t) just as residual-based

control charts do. The difference is that in residual-based control charts, the chart statistic depends on the model, and the control limits do not. In control charts applied directly to the autocorrelated process data, the control limits depend on the model, whereas the chart statistic does not. If the estimated model is inaccurate in either case, the control limits will fail to provide the desired ARL.

Two main objectives exist in this dissertation. The first objective is to investigate the sensitivity of parameter modeling errors on the foregoing SPC procedures that deal with autocorrelated processes. The measure of sensitivity is derived in an analytical form. Thus, the sensitivity is quantified and used for comparing the robustness between the methods. The second objective is to develop robust design procedures for the SPC with respect to parameter modeling errors from sensitivity results. Two kinds of robust design methods are introduced. The first approach for robust design uses the concept of worst-case scenario to guarantee a desired level of control specification. The next robust design method uses the expected variance of chart statistic. The resulting control limits are widened by an amount that depends on a number of factors, including the level of model uncertainty. Throughout the dissertation, the EWMA filter is used as a chart statistic.

I.2 Relation to Prior Work

The majority of the research in the SPC has focused on evaluating performances of various SPC techniques or comparing performances between SPC techniques under the assumption of given true parameters. There are much less results of unknown

parameters when compared to the results of known parameters. Even so, there have been more results of independent process data than autocorrelated process data because of tractable number of unknown parameters, which are usually a process mean and variance. Most work has investigated estimation effects of the process mean and variance on performances of SPC charts within the assumption of iid processes.

Ghosh, Reynolds and Van Hui (1981) studied the effect of unknown process variance when the $\overline{X}$ chart is used to monitor a process mean. Quesenberry (1993) investigated the effects of sample size on the run length distribution of control charts. The author recommended the sample size for Shewhart and $\overline{X}$ charts based on empirical evidence. Chen (1997) enhanced the result of $\overline{X}$ chart by the analytical derivation of the run length distribution for three different estimation methods of the process variance. Jones, Champ and Rigdon (2001) investigated the run length distribution of estimated process parameters in implementing EWMA control charts. They suggested the use of bigger sample size with small values of λ, the EWMA constant, when the estimated process parameters were applied. Jones (2002) suggested using widened control limits to assure the desired level of in-control ARL, consequently to reduce false alarm rate for the EWMA control charts. Also the author gave the values of constant L that provides a desired in-control ARL with various combinations of sample size and in-control ARL magnitude.

There have been limited efforts to understand the effects of parameter estimation errors in using control charts for autocorrelated processes. As a sensitivity analysis purpose, some empirical results were shown. Adams and Tseng (1998) empirically

investigated the sensitivity of four kinds of control charts when residuals were extracted from the autocorrelated process data assumed to follow an autoregressive (AR)(1) or integrated moving average (IMA)(1,1) model. According to the direction of estimation errors, the performances of residual-based control charts were severely affected. Especially, the resulting in-control ARLs of the EWMA and CUSUM control charts were drastically decreased than the aimed in-control ARL when the parameters were estimated such that the residuals are positively autocorrelated. This is because of the structure property of EWMA and CUSUM statistics that use weighted past values.

Apley and Shi (1999) also investigated the effect of modeling errors for three residual-based control charts — generalized likelihood ratio test (GLRT), CUSUM and Shewhart individual — using simulations. They showed that model estimation errors caused large and adverse impact on performances of the control charts. Lu and Reynolds (1999) also studied the robustness of the SPC chart with the empirical evidence. They concluded that the performance of control charts on the residuals or the original autocorrelated data was strongly influenced by the model estimation errors. These are, however, not well suitable for robust control chart design purposes.

As a robust design procedure, Apley (2002) proposed a design method to be robust to parameter modeling errors for autocorrelated processes. The author represented the variance of EWMA with a first-order of Taylor approximation and provided closed forms of the EWMA variance for first order ARMA processes. This method was the pioneering work in the robust SPC design for autocorrelated data. The proposed

control limits were widened properly in accordance with the level of model uncertainty to guard the aimed in-control ARL.

Previous studies on the robustness of SPC charts for autocorrelated processes have been mainly based on empirical methods and thus analytical results are practically nonexistent. In addition, these are mainly for sensitivity analysis purposes and are not well suited for robust control chart design purposes. In this dissertation, the sensitivity analysis of SPC charts will be investigated using analytical methods. In addition to providing better insight into the reasons for lack of robustness, the analytical results will be used to develop two different approaches for designing robust SPC control charts for autocorrelated processes.

I.3 Outline of the Dissertation

Chapter II derives a general result for the sensitivity of the variance of a linear-filtered ARMA(p,q) process and discusses the sensitivity and robustness of EWMAs on x_t and on e_t as a special case of the general result. Since the sensitivity of a control chart is closely related with the sensitivity of the variance of the chart statistic, the sensitivity is quantified by the sensitivity of the variance of chart statistic with respect to ARMA parameter estimation errors. Therefore, the measure of sensitivity is represented by partial derivatives of the variance at each estimated parameter. Finally, the measure is formed by the weighted sum of autocorrelation function. If an ARMA process is determined, then the magnitude of the sensitivity is determined based on the autocorrelation function of the process.

One conclusion is that residual-based EWMA control chart is no less robust than EWMA control chart on x_t. This is significant because in much of the criticism of residual-based charts, it was implied that alternative methods (i.e., control charts applied directly to x_t) would be more robust. Also, we discuss the sensitivity of control charts applied to processes in which autocorrelation is removed via feedback control. The sensitivity result of the residual-based control chart is identical to that of control chart on the feedback-controlled output when minimum variance (MV) controller is employed to remove the autocorrelation. Although the focus is on the analytical sensitivity analysis, some empirical results, which combine performance and sensitivity information, are also provided.

Chapter III proposes a robust design method for residual-based control charts from the result of Chapter II. To account for uncertainty in the estimated parameters and guard against a situation in which the in-control ARL is substantially shorter than desired, a reasonable precaution is to use control limits that are wider than those used when the model is assumed perfect. This chapter presents a method for systematically widening the control limits based on the "worst-case" design approach.

Considering the uncertainty in the true parameters, a confidence interval for the standard deviation of chart statistic is approximated. To find an approximate confidence interval, a first-order Taylor approximation is used. Then, the upper boundary of the confidence interval can be viewed as a worst-case (maximum) value for the true standard deviation of chart statistic and the worst-case standard deviation is used for the control chart design. Decreased power by widened control limits is

indicated as a weakness and comparison results to Shewhart chart, which is least sensitive to parameter modeling errors, is discussed. Also, the sample size requirements are investigated.

Chapter IV develops another robust design procedure for residual-based control charts. When modeling errors exist, the actual variance of a chart statistic will be different from the ideal variance that assumes no modeling errors. The proposed design approach also quantifies the differences between the actual and ideal variances and modifies the control limits accordingly. The actual variance of the chart statistic is represented using a second-order Taylor approximation in this method. After taking the expectation of the second-order approximation, with respect to the parameter uncertainty, the result is an expression for the expected variance as a function of parameter estimates and their covariances.

To evaluate the proposed method, it is compared to an existing robust design method and the robust design method in Chapter III. From comparison results, this proposed approach achieves a more suitable balance between false alarms and control chart power.

Chapter V summarizes conclusions and discusses future work.

CHAPTER II

SENSITIVITY AND ROBUSTNESS OF EWMA CHARTS FOR

AUTOCORRELATED PROCESSES WITH MODELING ERRORS

II.1 Introduction

The EWMA statistic on x_t, which is assumed to follow (1.1), is calculated recursively via $z_t = (1-\lambda)z_{t-1} + \lambda x_t$, where $0 < \lambda \le 1$ is the EWMA parameter. The control chart signals a mean shift if z_t falls outside the upper and lower control limits (Zhang 1998)

$$\{LCL, UCL\} = \pm L\sigma_z,$$

where σ_z is the (steady-state) standard deviation of z_t, and the constant L can be chosen to provide a specified false alarm rate assuming estimated model is perfect. Zhang (1998) provided a straightforward approach for calculating σ_z as a function of the autocorrelation of x_t.

The standard residual-based EWMA design is to set $\pm L\sigma_z$ control limits on the EWMA of the form $z_t = (1-\lambda)z_{t-1} + \lambda e_t$, where $e_t = \hat{\Theta}^{-1}(B)\hat{\phi}(B)x_t$ and the "^" symbol denotes an estimate of a quantity (Lu and Reynolds 1999). If there are no modeling errors, then $\sigma_z = \sigma_a \lambda^{1/2}(2-\lambda)^{-1/2}$, and L can be chosen to provide a desired false alarm rate or in-control average run length (ARL) (Montgomery 2001; Lucas and Saccucci 1990).

In this chapter we derive relatively compact, closed-form analytical expressions for the sensitivity of the residual-based EWMA variance σ_z^2 with respect to ARMA parameter estimation errors. We also derive analogous sensitivity results for an EWMA applied directly to the autocorrelated x_t. The sensitivity results that we develop in this chapter will be also used in Chapters III and IV in order to suitably widen the control limits of a residual-based EWMA by taking into account the level of model parameter uncertainty. Standard results for the covariance matrix of ARMA parameter estimates will be used to quantify the uncertainty.

The format of the remainder of the chapter is as follows. In Chapter II.2 we derive a general result for the sensitivity of the variance of a linear-filtered ARMA(p,q) process. In Chapter II.3, we derive sensitivity results for the residual-based EWMA as a special case of the general result. Although the results for a residual-based EWMA are of simpler form than for an EWMA on x_t, the results for x_t are relatively simple for first-order ARMA processes. These are derived in Chapter II. 4. Chapter II.5 provides a discussion on the sensitivity of control charts applied to processes in which autocorrelation is removed via feedback-control. In Chapter II.6, we compare the sensitivity and performance of a residual-based EWMA versus an EWMA on x_t.

II.2 General Sensitivity Results for Linear-filtered ARMA Processes

Consider the output z_t of a linear filter $H(B) = \sum_{j=0}^{\infty} h_j B^j$ applied to an ARMA(p,q) process x_t, where $\{h_j: j = 0, 1, 2, \ldots\}$ are the impulse response coefficients (Box, et al. 1994) of $H(B)$. Write this as

$$z_t = H(B)x_t = G(B)a_t = \sum_{j=0}^{\infty} g_j a_{t-j} \tag{2.1}$$

where $G(B) = \Phi^{-1}(B)\Theta(B)H(B) = \sum_{j=0}^{\infty} g_j B^j$. In this chapter, we derive a general result for the sensitivity of σ_z^2 with respect to the ARMA parameters. Note that the EWMAs on x_t and on e_t are special cases with $H(B) = (1-v)(1-vB)^{-1}$ and $H(B) = (1-v)(1-vB)^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)$, respectively, where $v = 1-\lambda$. The variance of z_t is given by (Box, et al. 1994)

$$\sigma_z^2 = \sigma_a^2 \sum_{j=0}^{\infty} g_j^2 \tag{2.2}$$

Suppose that we have the estimates $\hat{\Theta}(B)$ and $\hat{\Phi}(B)$ available, and that the filter $H(B)$ does not depend on the unknown, true parameters $\Theta(B)$ and $\Phi(B)$. $G(B)$ and σ_z^2 are unknown because they depend on $\Phi(B)$ and $\Theta(B)$, but consider $\hat{G}(B) = \hat{\Phi}^{-1}(B)\hat{\Theta}(B)H(B)$ and

$$\hat{\sigma}_z^2 = \sigma_z^2\Big|_{\gamma=\hat{\gamma}} = \sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j^2 \tag{2.3}$$

where $\gamma = [\phi_1 \ \phi_2 \ ... \ \phi_p \ \theta_1 \ \theta_2 \ ... \ \theta_q]^T$ is the vector of ARMA parameters. As measures

of the sensitivity of σ_z^2 with respect to the ARMA parameters, we use the quantities

$$S(\phi_i) = \frac{\left.\frac{\partial \sigma_z^2}{\partial \phi_i}\right|_{\gamma=\hat{\gamma}}}{\hat{\sigma}_z^2} \quad : i = 1, 2, \ldots, p, \text{ and} \tag{2.4}$$

$$S(\theta_i) = \frac{\left.\frac{\partial \sigma_z^2}{\partial \theta_i}\right|_{\gamma=\hat{\gamma}}}{\hat{\sigma}_z^2} \quad : i = 1, 2, \ldots, q. \tag{2.5}$$

Theorem: For a general linear filter $z_t = H(B)x_t$, the sensitivities are

$$S(\phi_i) = 2 \sum_{k=0}^{\infty} P_k \rho_{i+k} \quad : i = 1, 2, \ldots, p, \text{ and} \tag{2.6}$$

$$S(\theta_i) = -2 \sum_{k=0}^{\infty} Q_k \rho_{i+k} \quad : i = 1, 2, \ldots, q, \tag{2.7}$$

where ρ_j denotes the autocorrelation function of $z_t|_{\gamma=\hat{\gamma}}$ (z_t when there are no

modeling errors), and $\{P_j : j = 0, 1, 2, \ldots\}$ and $\{Q_j : j = 0, 1, 2, \ldots\}$ denote the impulse

response coefficients of $P(B) = \hat{\Phi}^{-1}(B) = \sum_{j=0}^{\infty} P_j B^j$ and

$Q(B) = \hat{\Theta}^{-1}(B) = \sum_{j=0}^{\infty} Q_j B^j$, respectively.

Proof: Differentiating (2.2) gives

$$\left.\frac{\partial \sigma_z^2}{\partial \phi_i}\right|_{\gamma=\hat{\gamma}} = 2\sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j d_j^{\phi_i} \text{, and} \tag{2.8}$$

$$\left.\frac{\partial \sigma_z^2}{\partial \theta_i}\right|_{\gamma=\hat{\gamma}} = 2\sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j d_j^{\theta_i} \tag{2.9}$$

where $d_j^{\phi_i} = \partial g_j/\partial \phi_i|_{\gamma=\hat{\gamma}}$ and $d_j^{\theta_i} = \partial g_j/\partial \theta_i|_{\gamma=\hat{\gamma}}$. From the relationship $G(B) = \Phi^{-1}(B)\Theta(B)H(B)$, we have $g_j - \phi_1 g_{j-1} - \phi_2 g_{j-2} - \dots - \phi_p g_{j-p} = h_j - \theta_1 h_{j-1} - \theta_2 h_{j-2} - \dots - \theta_q h_{j-q}$. Differentiating both sides with respect to ϕ_i and θ_i, and evaluating the result at the ARMA parameter estimates gives

$$d_j^{\phi_i} - \hat{\phi}_1 d_{j-1}^{\phi_i} - \dots - \hat{\phi}_p d_{j-p}^{\phi_i} - \hat{g}_{j-i} = 0, \text{ and} \tag{2.10}$$

$$d_j^{\theta_i} - \hat{\phi}_1 d_{j-1}^{\theta_i} - \dots - \hat{\phi}_p d_{j-p}^{\theta_i} = -h_{j-i}, \tag{2.11}$$

where it is understood that $g_j = h_j = \hat{g}_j = d_j^{\phi_i} = d_j^{\theta_i} = 0$ for $j < 0$, and we have used the fact that $H(B)$ does not depend on the true ARMA parameters. If we view $d_j^{\phi_i}$ and $d_j^{\theta_i}$ as sequences in the index j, rearranging (2.10) and (2.11) gives

$$d_j^{\phi_i} = \hat{\Phi}^{-1}(B)\hat{g}_{j-i} = \sum_{k=0}^{\infty} P_k \hat{g}_{j-i-k}, \text{ and}$$

$$d_j^{\theta_i} = -\hat{\Phi}^{-1}(B)h_{j-i} = -\hat{\Phi}^{-1}(B)\hat{\Theta}^{-1}(B)\hat{\Phi}(B)\hat{g}_{j-i} = -\hat{\Theta}^{-1}(B)\hat{g}_{j-i} = -\sum_{k=0}^{\infty} Q_k \hat{g}_{j-i-k}$$

Using these in (2.4) and (2.5), the sensitivity measures become

$$S(\phi_i) = \frac{2\sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j \sum_{k=0}^{\infty} P_k \hat{g}_{j-i-k}}{\sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j^2} = 2 \sum_{k=0}^{\infty} P_k \frac{\sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j \hat{g}_{j-(i+k)}}{\sigma_a^2 \sum_{j=0}^{\infty} \hat{g}_j^2}, \text{ and}$$

$$S(\theta_i) = \frac{-2\sigma_a^2 \sum\limits_{j=0}^{\infty} \hat{g}_j \sum\limits_{k=0}^{\infty} Q_k \hat{g}_{j-i-k}}{\sigma_a^2 \sum\limits_{j=0}^{\infty} \hat{g}_j^2} = -2 \sum\limits_{k=0}^{\infty} Q_k \frac{\sigma_a^2 \sum\limits_{j=0}^{\infty} \hat{g}_j \hat{g}_{j-(i+k)}}{\sigma_a^2 \sum\limits_{j=0}^{\infty} \hat{g}_j^2} .$$

Recognizing that the denominator and numerator in the far right expressions are the variance and lag–$(k+i)$ autocovariance of $z_t |_{\gamma=\hat{\gamma}} = \sum_{j=0}^{\infty} \hat{g}_j a_{t-j}$ (Box, et al. 1994) completes the proof. ∎

$S(\phi_i)$ and $S(\theta_i)$ are the weighted sums of the impulse response coefficients of $\hat{\Phi}^{-1}(B)$ and $\hat{\Theta}^{-1}(B)$, where the weights are given by the autocorrelation function of $z_t |_{\gamma=\hat{\gamma}}$. For the case that z_t is an EWMA, an EWMA on positively autocorrelated data x_t will have a more slowly decaying autocorrelation function than an EWMA on e_t with the same value of λ. Consequently, it will generally be the case that an EWMA on x_t is more sensitive to modeling errors than an EWMA on e_t. We demonstrate this more concretely in Chapter II.4 for the special case of first-order ARMA processes. This conclusion is somewhat surprising given that an EWMA on the ARMA residuals may appear to rely more heavily on the ARMA model than an EWMA on x_t. The control limits ($\pm L\sigma_z$) for an EWMA on x_t do, however, depend heavily on the ARMA parameters.

II.3 Sensitivity Results for the Residual-based EWMA

The sensitivity expressions in the theorem of the preceding chapter simplify considerably when z_t is a residual-based EWMA. In this case, $H(B) = (1-v)(1-vB)^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)$, and $\hat{G}(B) = \hat{\Phi}^{-1}(B)\hat{\Theta}(B)H(B) = (1-v)(1-vB)^{-1}$, where $v = 1-\lambda$. With no modeling errors, the residual-based EWMA is the first-order AR process $z_t|_{\gamma=\hat{\gamma}} = (1-v)(1-vB)^{-1}a_t$, with autocorrelation function $\rho_j = v^j$. Substituting this into (2.6) and (2.7), the sensitivities for the residual-based EWMA become

$$S_e(\phi_i) = 2\sum_{k=0}^{\infty} P_k\, v^{k+i} = 2v^i \sum_{k=0}^{\infty} P_k\, B^k|_{B=v} = 2v^i \hat{\Phi}^{-1}(B)|_{B=v} = \frac{2v^i}{\hat{\Phi}(v)} \qquad (2.12)$$

and

$$S_e(\theta_i) = -2\sum_{k=0}^{\infty} Q_k\, v^{k+i} = -2v^i \sum_{k=0}^{\infty} Q_k\, B^k|_{B=v} = -2v^i \hat{\Theta}^{-1}(B)|_{B=v} = \frac{-2v^i}{\hat{\Theta}(v)} \qquad (2.13)$$

where $\hat{\Phi}(v) = 1 - \hat{\phi}_1 v - \hat{\phi}_2 v^2 - \ldots - \hat{\phi}_p v^p$, and $\hat{\Theta}(v) = 1 - \hat{\theta}_1 v - \hat{\theta}_2 v^2 - \ldots - \hat{\theta}_q v^q$. We have added the subscript e on the sensitivities to indicate they are for an EWMA on e_t.

$S_e(\phi_i)$ and $S_e(\theta_i)$ have clearer interpretations if we factor the AR and MA polynomials in terms of their roots. Consider the factorization of $\hat{\Phi}(B) = (1-\eta_1 B)(1-\eta_2 B)\ldots(1-\eta_p B)$ in terms of its roots $\{\eta_1, \eta_1, \ldots, \eta_p\}$. The magnitude of the sensitivity in (2.12) becomes

$$\left|S_e(\phi_i)\right| = 2v^i \prod_{j=1}^{p} \frac{1}{|1-\eta_j v|} \qquad (2.14)$$

Each term $|1-\eta_j \nu|$ in the denominator in (2.14) represents the distance between the scaled root $\eta_j \nu$ and the point 1.0 (i.e., the intersection of the unit circle and the real axis) in the complex plane. This is illustrated in Figure 1 for a complex root and $\nu =$ 0.9. The sensitivity will be large if any root is close to the point 1.0 and the EWMA parameter λ is small (ν close to 1). Complex conjugate roots near the stability boundary (the unit circle) do not necessarily result in large sensitivity. In contrast, roots on the positive real axis near the stability boundary always result in large sensitivity if λ is small. Similar results hold for $S_e(\theta_i)$ in terms of the roots of $\hat{\Theta}(B)$.
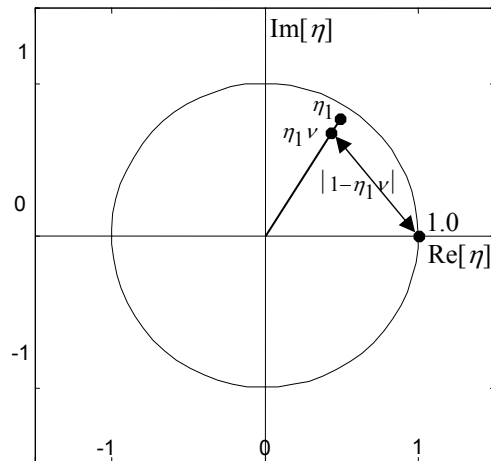
Figure 1 Illustration of the distance between the scaled root $\eta_1 \nu$ and the point 1.0 in the complex plane. Smaller distances increase the sensitivity $S_e(\phi_i)$.

II.4 Sensitivity Results for First-order ARMA Processes

Although there are no simple closed-form expressions analogous to (2.12) and (2.13) for the sensitivity of the EWMA on x_t for general ARMA processes, (2.6) and (2.7) can be simplified for ARMA(1,1) processes. When x_t is ARMA(1,1), the EWMA statistic $z_t|_{\gamma=\hat{\gamma}}$ with no modeling errors follows the ARMA(2,1) model

$$z_t|_{\gamma=\hat{\gamma}} = \frac{(1-v)(1-\hat{\theta}B)}{(1-vB)(1-\hat{\phi}B)} a_t \tag{2.15}$$

where we have dropped the subscripts on the ARMA parameters. A closed-form expression for the autocorrelation function of the ARMA(2,1) process is (Pandit and Wu 1983)

$$\rho_k = \frac{1}{c_1 + c_2}\left(c_1 \hat{\phi}^k + c_2 v^k\right) \tag{2.16}$$

where

$$c_1 = \frac{(\hat{\phi}-\theta)(1-\hat{\phi}\theta)}{(\hat{\phi}-v)(1-\hat{\phi}v)(1-\hat{\phi}^2)},$$

$$c_2 = \frac{-(v-\theta)(1-\theta v)}{(\hat{\phi}-v)(1-\hat{\phi}v)(1-v^2)}, \text{ and}$$

$$c_1 + c_2 = \frac{1}{(\hat{\phi}-v)(1-\hat{\phi}v)}\left[\frac{(\hat{\phi}-\theta)(1-\hat{\phi}\theta)}{1-\hat{\phi}^2} - \frac{(v-\theta)(1-\theta v)}{1-v^2}\right] = \frac{(1+\theta^2)(1+\hat{\phi}v)-2\theta(v+\hat{\phi})}{(1-\hat{\phi}^2)(1-v^2)(1-\hat{\phi}v)}$$

From Box, et al. (1994), the impulse response coefficients of $P(B) = (1-\hat{\phi}B)^{-1}$ and $Q(B) = (1-\hat{\theta}B)^{-1}$ are $P_k = \hat{\phi}^k$ and $Q_k = \hat{\theta}^k$. Substituting these and (2.16) into (2.6) gives the following expressions for the sensitivity of the EWMA on x_t.

$$S_x(\phi) = 2 \sum_{k=0}^{\infty} P_k \rho_{1+k} = \frac{2}{c_1 + c_2} \sum_{k=0}^{\infty} \hat{\phi}^k \left(c_1 \hat{\phi}^{k+1} + c_2 v^{k+1} \right) = \frac{2}{c_1 + c_2} \left[\frac{c_1 \hat{\phi}}{1 - \hat{\phi}^2} + \frac{c_2 v}{1 - \hat{\phi} v} \right]$$

$$= \frac{2}{c_1 + c_2} \left[\frac{v(c_1 + c_2)}{1 - \hat{\phi} v} + \frac{c_1 \hat{\phi}}{1 - \hat{\phi}^2} - \frac{c_1 v}{1 - \hat{\phi} v} \right] = \frac{2}{c_1 + c_2} \left[\frac{v(c_1 + c_2)}{1 - \hat{\phi} v} + \frac{c_1 (\hat{\phi} - v)}{(1 - \hat{\phi}^2)(1 - \hat{\phi} v)} \right]$$

$$= \frac{2v}{1 - \hat{\phi} v} + \frac{2(\hat{\phi} - \hat{\theta})(1 - \hat{\phi}\hat{\theta})(1 - v^2)}{(1 - \hat{\phi}^2)(1 - \hat{\phi} v)\left[(1 + \hat{\theta}^2)(1 + \hat{\phi} v) - 2\hat{\theta}(v + \hat{\phi})\right]} \tag{2.17}$$

and

$$S_x(\theta) = -2 \sum_{k=0}^{\infty} Q_k \rho_{1+k} = \frac{-2}{c_1 + c_2} \sum_{k=0}^{\infty} \hat{\theta}^k \left(c_1 \hat{\phi}^{k+1} + c_2 v^{k+1} \right) = \frac{-2}{c_1 + c_2} \left[\frac{c_1 \hat{\phi}}{1 - \hat{\phi}\hat{\theta}} + \frac{c_2 v}{1 - \hat{\theta} v} \right]$$

$$= \frac{-2}{c_1 + c_2} \left[\frac{v(c_1 + c_2)}{1 - \hat{\theta} v} + \frac{c_1 \hat{\phi}}{1 - \hat{\phi}\hat{\theta}} - \frac{c_1 v}{1 - \hat{\theta} v} \right] = \frac{-2}{c_1 + c_2} \left[\frac{v(c_1 + c_2)}{1 - \hat{\theta} v} + \frac{c_1 (\hat{\phi} - v)}{(1 - \hat{\phi}\hat{\theta})(1 - \hat{\theta} v)} \right]$$

$$= \frac{-2v}{1 - \hat{\theta} v} - \frac{2(\hat{\phi} - \hat{\theta})(1 - v^2)}{(1 - \hat{\theta} v)\left[(1 + \hat{\theta}^2)(1 + \hat{\phi} v) - 2\hat{\theta}(v + \hat{\phi})\right]} \tag{2.18}$$

In comparison, for the residual-based EWMA, (2.12) and (2.13) reduce to

$$S_e(\phi) = \frac{2v}{1 - \hat{\phi} v}, \text{ and} \tag{2.19}$$

$$S_e(\theta) = \frac{-2v}{1 - \hat{\theta} v} \tag{2.20}$$

which are identical to the first terms in (2.17) and (2.18). Note that $S_x(\phi)$ for an AR(1)

process and $S_x(\theta)$ for an MA(1) process are obtained by substituting $\hat{\theta} = 0$ into (2.17)

and $\hat{\phi} = 0$ into (2.18), respectively.

For ARMA(1,1) processes we can restrict attention to the case $|\hat{\phi}| < 1$ and $|\hat{\theta}| < 1$, which must hold for stable, invertible ARMA processes. Moreover, the EWMA parameter is restricted to $0 \le v < 1$. Because negatively autocorrelated data are rare in industrial processes, we might also restrict attention to the case that $\hat{\theta} < \hat{\phi}$. Under these conditions, it is straightforward to show that $|S_x(\phi)| > |S_e(\phi)|$ and $|S_x(\theta)| > |S_e(\theta)|$ always hold. By inspection of (2.17) and (2.18), we only need to show that

$$\left(1 + \hat{\theta}^2\right)\left(1 + \hat{\phi}v\right) - 2\hat{\theta}\left(v + \hat{\phi}\right) > 0, \text{ and this follows because } 1 + \hat{\phi}v > \left|v + \hat{\phi}\right| \text{ and}$$

$$1 + \hat{\theta}^2 > \left|2\hat{\theta}\right|.$$

Example. As an example, consider the Series A data from Box, et al. (1994), which are 197 concentration measurements from a chemical process. Box, et al. (1994) found that an ARMA(1,1) model fit the data well, and the estimated parameters were $\hat{\phi} = 0.87$, $\hat{\theta} = 0.48$, and $\hat{\sigma}_a^2 = 0.098$. Suppose we intend to monitor the process for mean shifts using EWMA control charts on x_t and e_t with EWMA parameter $\lambda = 0.10$. If we neglect modeling errors, the assumed standard deviation for the EWMA on x_t is $\hat{\sigma}_{z,x} = 0.220$, which follows from (2.4) using the impulse response coefficients of the ARMA(2,1) process (2.15). The assumed standard deviation for the EWMA on e_t is $\hat{\sigma}_{z,e} = \hat{\sigma}_a (1-v)^{1/2}(1+v)^{-1/2} = 0.0718$. In order to gage the effects of modeling errors on the EWMAs, we can calculate the sensitivities. For the residual-based EWMA, (2.19) and (2.20) give $S_e(\phi) = 8.29$ and $S_e(\theta) = -3.17$. For the EWMA on x_t, (2.17)

and (2.18) give $S_x(\phi) = 11.60$ and $S_x(\theta) = -3.70$. Both charts are sensitive to modeling

errors, although the EWMA on x_t is somewhat more sensitive. To illustrate the effects

of modeling errors, suppose that θ and σ_a coincide with their estimates, but that $\phi =$

0.90. The top two panels in Figure 2 show 1000 simulated observations of x_t and e_t,

and the bottom two panels show the EWMAs on x_t and e_t. The control limits for the

EWMA on x_t and e_t. were set at $\pm 3\,\hat{\sigma}_{z,x} = \pm 0.660$ and $\pm 3\,\hat{\sigma}_{z,e} = \pm 0.216$, respectively.

No mean shifts were added, so that the frequent alarms in the control chart are all false

alarms.

Because ϕ differs from $\hat{\phi}$, the residuals actually follow the ARMA(1,1) model

$$e_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}x_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)}\frac{\Theta(B)}{\Phi(B)}a_t = \frac{1-0.87B}{1-0.90B}a_t. \tag{2.21}$$

With ϕ underestimated, the residual autocorrelation is positive. Although the standard

deviation of the residuals in (2.21) is only 0.24% larger than σ_a, and the residual

autocorrelation at any given lag is quite small, the autocorrelation dies out slowly. The

result is that the actual standard deviation $\sigma_{z,e}$ of the EWMA on the residuals is

substantially larger than $\hat{\sigma}_{z,e}$. Similar arguments hold for the EWMA on x_t, which

also has an inflated standard deviation. The sensitivities can be used to approximate

the increase in the EWMA standard deviation. For the EWMA on x_t, the approximate

percentage increase in the EWMA variance is $S_x(\phi)(\phi - \hat{\phi}) = 34.8\%$. Thus, the

approximate percentage increase in the EWMA standard deviation is $(1.348)^{1/2} - 1 =$
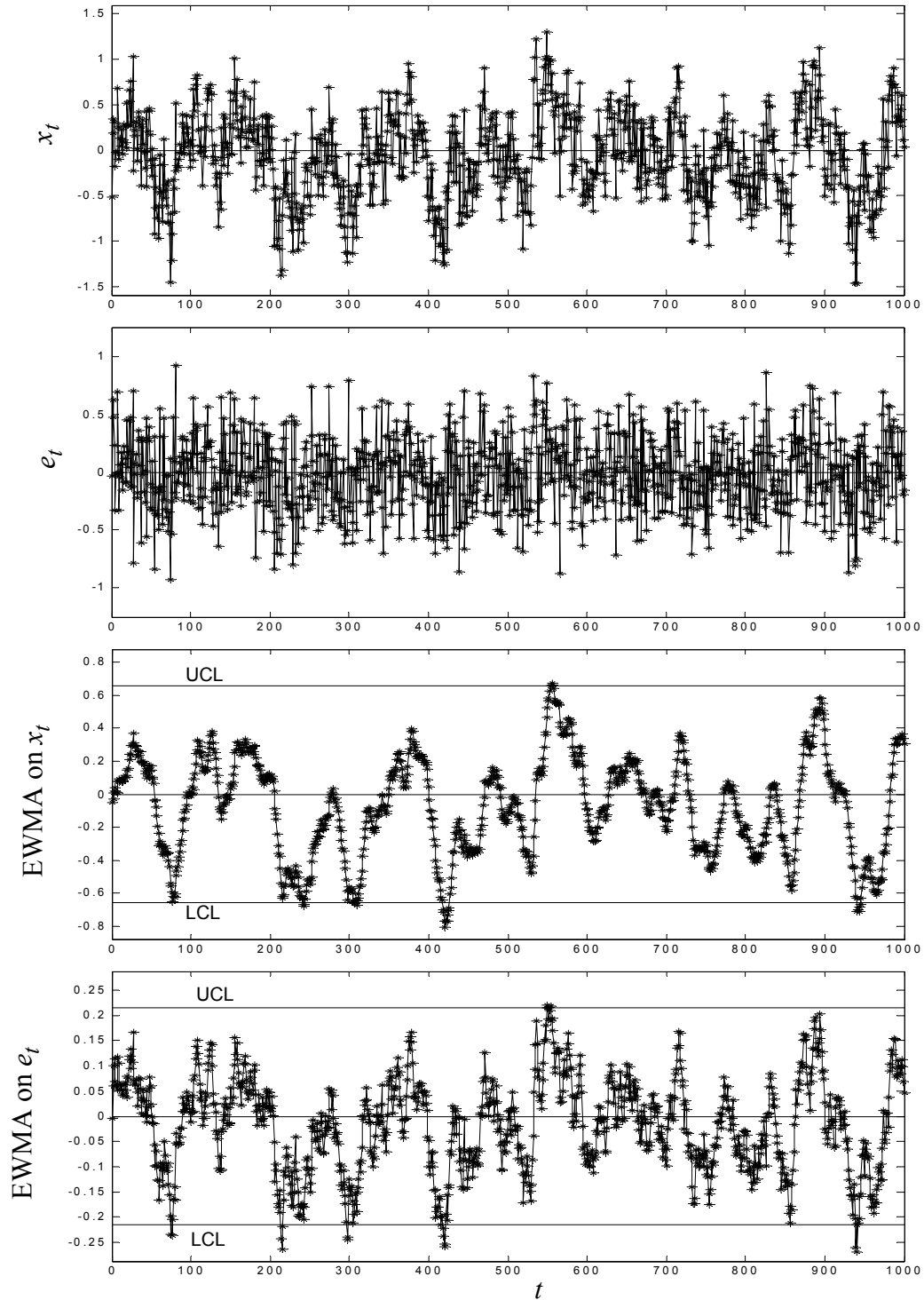
Figure 2 Example illustrating the increased false alarm rate for EWMA charts with modeling errors. The four panels from top to bottom are x_t, e_t, an EWMA on x_t, and an EWMA on e_t.

15.8%. Similarly, the approximate percentage increase in the standard deviation of the residual-based EWMA is 11.8%. The inflated EWMA standard deviation is evident in the frequent false alarms in Figure 2.

Although the sensitivities provide a reasonable basis for comparison, they result in a first-order approximation of the effects of modeling errors on the EWMA standard deviation. The exact effects can be calculated using (2.2) and the impulse response coefficients for z_t with parameter errors. Based on this it can be shown that the actual standard deviation of the EWMA on x_t is $\sigma_{z,x} = 0.267$, which is 21.3% larger than $\hat{\sigma}_{z,x}$. Similarly, the actual standard deviation of the residual-based EWMA is $\sigma_{z,e} = 0.0828$, which is 15.3% larger than $\hat{\sigma}_{z,e}$. Because $L = 3$ was used in the control limits, the assumed false alarm probability for both charts is 0.0027. For the EWMA on x_t, for which $\sigma_{z,x}$ is 21.3% larger than $\hat{\sigma}_{z,x}$, the actual false alarm probability is 0.0134 – roughly five times larger than the assumed value. For the EWMA on e_t, for which $\sigma_{z,e}$ is 15.3% larger than $\hat{\sigma}_{z,e}$, the actual false alarm probability is 0.0093.

Although it is tempting to conclude from the preceding discussion that the residual-based EWMA is more robust than an EWMA on x_t, the direct comparison is not entirely fair. In the preceding discussion, the two charts were compared under the assumption that the same value for the EWMA parameter was selected for both. If x_t has large positive autocorrelation, to achieve more comparable performance in detecting mean shifts one might use a smaller value of v for an EWMA on x_t than for

a residual-based EWMA. As v decreases, the EWMA on x_t becomes less sensitive to modeling errors. In Chapter II.6 we provide a more elaborate comparison of the EWMAs on x_t and on e_t in which we consider both performance and sensitivity.

II.5 Removing Autocorrelation with Feedback Control

In light of the lack of robustness of control charts for autocorrelated data, some authors (e.g, Adams and Tseng 1998) have recommended removing autocorrelation via feedback control when applicable and applying the control charts to the closed-loop output. To illustrate, suppose that the process output x_t obeys the model $x_t = \beta u_{t-1} + d_t$, where u_t is an adjustable process input, β represents the effects of the input on the output, $d_t = \Phi^{-1}(B)\Theta(B)a_t$ is an ARMA process disturbance, and the output target value is zero. Refer to Figure 3. It is well known (Box et al. 1994) that if minimum variance control is used and there is no model uncertainty, then the closed-loop process output is $x_t = a_t$. Consequently, the closed-loop output is uncorrelated, and traditional control charts can be applied.

To understand the effects of parameter uncertainty on the closed-loop output, write the minimum variance control law as (Åström and Wittenmark 1990)

$$u_t = \frac{-\hat{H}(B)}{\hat{\beta}\hat{\Phi}(B)} x_t ,$$

where $B\hat{H}(B) = \hat{\Theta}(B) - \hat{\Phi}(B)$. If we temporarily assume $\beta = \hat{\beta}$, then substituting the control law into the model $x_t = \beta u_{t-1} + \Phi^{-1}(B)\Theta(B)a_t$ gives

$$\left(1 + \frac{B\hat{H}(B)}{\hat{\Phi}(B)}\right)x_t = \frac{\Theta(B)}{\Phi(B)}a_t$$

Substituting $B\hat{H}(B) = \hat{\Theta}(B) - \hat{\Phi}(B)$ and rearranging terms, it follows that the closed-loop process output with ARMA parameter errors obeys the model

$$x_t = \frac{\hat{\Phi}(B)\Theta(B)}{\hat{\Theta}(B)\Phi(B)}a_t$$

Because this is precisely the equation describing the residuals of the ARMA disturbance model, we see that ARMA parameter errors would affect a control chart applied to the closed-loop output in exactly the same manner as a residual-based control chart. If in addition we consider errors in the parameter β, it is reasonable to conclude that control charts applied to the closed-loop output would be even less robust than residual-based control charts.
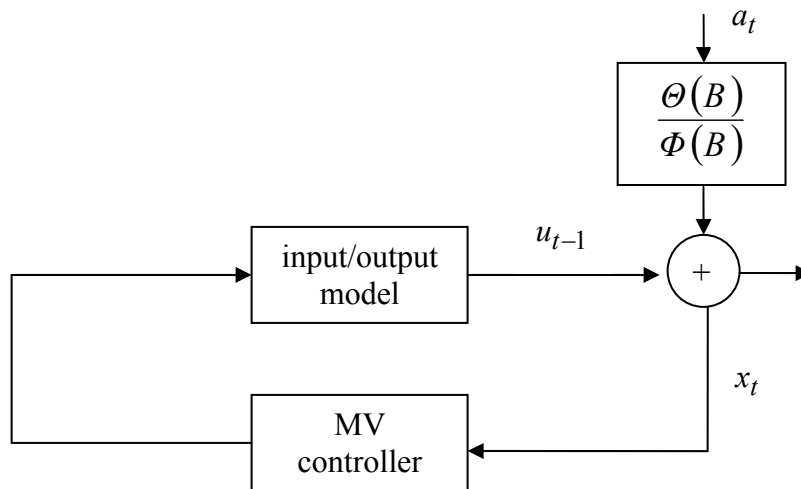
Figure 3 Block diagram of minimum variance controlled process.

II.6 Performance and Sensitivity Comparison

As discussed in Chapters II.3 and II.4, a residual-based EWMA is generally more robust than an EWMA on x_t if both charts use the same value of λ. This is not necessarily an appropriate basis for comparison, however. The value of λ that would typically be selected for each chart depends on many factors, including the out-of-control ARL performance of the chart. One might compare the two charts by choosing each λ so that the out-of-control ARL with no modeling errors are equivalent for the two charts, and then comparing the sensitivities of each. However, it will not always be possible to equate the out-of-control ARLs for the two charts. Consequently, we compare the two charts by selecting values of λ that provide equal sensitivities, and then comparing the resulting out-of-control ARL performance. All ARL values are for the case that there are no modeling errors and were calculated using Monte Carlo simulation with 10,000 replicates.

For each replicate, the a_t sequence was generated from the standard normal distribution, and then the x_t sequence was generated using (2.1). For the EWMA on x_t, the initial values of the x_t sequence were discarded, so that the remaining sequence can be assumed to have reached steady state. The EWMA was then applied directly to x_t with control limits chosen to provide an in-control ARL of 500 (with no modeling errors). To calculate an out-of-control ARL, a mean shift of magnitude μ was added to x_t at the initial timestep (but after the initial transient data was discarded). Similar procedures were used for the EWMA on e_t. After generating the residuals, the initial

values containing transient dynamics were discarded. The mean shift was added to x_t before generating the residuals, at the timestep corresponding to the first retained residual. The standard error for all ARL values was roughly 1%.

Tables 1 through 4 show the results of the Monte Carlo Simulation for four different ARMA models. ARL(μ) denotes the ARL for a mean shift of size $\mu\sigma_a$. The in-control ARL was 500 for all cases. Each column of the tables contains values for both EWMAs except for the cases where the last rows of the table can only display EWMA on e_t values, with the values for the EWMA on x_t in parentheses. The numbers in bold font in each of the ARL columns indicates the smallest out-of-control ARL for that size mean shift. Tables 1 and 2 are for AR(1) processes with ϕ=0.9 and ϕ=0.5, respectively. Each row compares the results for the two EWMAs with λ chosen to provide a common sensitivity $S(\phi)$ with respect to the AR parameter. For the cases where only EWMA on e_t values are displayed, there exist no values of λ that satisfies the common sensitivity for EWMA on x_t. Table 3 is for the ARMA(1,1) example considered in Chapter II.4, and Table 4 is for an ARMA(1,1) process with $\phi = 0.7$ and $\theta = 0.3$. In Tables 3 and 4, each row compares the two EWMAs with common AR parameter sensitivity. The resulting MA parameter sensitivies $S(\theta)$ are also shown.

Although Tables 1 through 4 indicate that the out-of-control ARL performances of the two charts are generally comparable for common sensitivity, the residual-based EWMA appears to have slightly better performance for most cases. Moreover, for each specific size mean shift, the minimum ARL (indicated by bold font) is generally

Table 1 Comparison of ARL performance for EWMAs on e_t and on x_t with common sensitivity and in-control ARL of 500. x_t is AR(1) with $\phi = 0.9$. The values for the EWMA on x_t are in parentheses.

$S(\phi)$	λ	L	ARL(0)	mean shift magnitude (in units of σ_a)				
				ARL(1)	ARL(2)	ARL(3)	ARL(4)	ARL(5)
18.0	0.0110 (0.0243)	2.0177 (1.9723)	500 (500)	**213** (**225**)	**90.1** (**92.9**)	51.6 (50.7)	33.6 (33.0)	24.1 (24.0)
17.0	0.0173 (0.0409)	2.2156 (2.1177)	500 (500)	225 (245)	91.6 (97.6)	50.1 (**50.5**)	31.9 (31.2)	21.7 (21.5)
16.0	0.0244 (0.0621)	2.3570 (2.2287)	500 (500)	231 (259)	93.7 (102)	**48.6** (50.9)	30.2 (30.2)	20.0 (19.9)
15.0	0.0323 (0.0901)	2.4636 (2.3148)	500 (500)	244 (270)	98.0 (109)	49.5 (52.6)	**29.6** (**29.8**)	18.9 (19.0)
14.0	0.0411 (0.1287)	2.5471 (2.3957)	500 (500)	260 (286)	105 (118)	52.0 (55.5)	30.0 (30.5)	18.4 (18.6)
13.0	0.0511 (0.1851)	2.6212 (2.4722)	500 (500)	274 (295)	111 (123)	54.1 (57.7)	30.3 (31.1)	17.9 (18.2)
12.0	0.0625 (0.2749)	2.6791 (2.5504)	500 (500)	283 (300)	118 (128)	56.0 (59.3)	30.8 (30.5)	17.6 (17.7)
11.0	0.0756 (0.4345)	2.7337 (2.6444)	500 (500)	295 (307)	127 (134)	60.8 (61.5)	31.6 (31.3)	17.2 (17.3)
10.0	0.0909 (0.7501)	2.7866 (2.7551)	500 (500)	306 (306)	137 (136)	64.3 (62.2)	33.3 (31.4)	17.1 (**16.7**)
8.0	0.1304	2.8749	500	336	161	77.3	38.3	15.4
6.0	0.1892	2.9522	500	360	187	93.1	41.6	14.3
4.0	0.2857	3.0135	500	397	233	113	43.1	10.9
2.0	0.4737	3.0658	500	425	282	138	42.0	**7.16**

slightly smaller for the residual-based EWMA, and the corresponding sensitivity is also slightly smaller. Especially for Table 1 where the rows have only the values for EWMA on e_t, there exist no values of λ that can make the EWMA on x_t as robust as the EWMA on e_t with larger λ values. In addition, EWMA on e_t has a much smaller ARL value than EWMA on x_t at column ARL(5). Therefore, EWMA on e_t has better performance and more robustness in this case. These ideal cases can be also found in Table 3.

Table 2 Comparison of ARL performance for EWMAs on e_t and on x_t with common sensitivity and in-control ARL of 500. x_t is AR(1) with $\phi = 0.5$. The values for the EWMA on x_t are in parentheses.

$S(\phi)$	λ	L	mean shift magnitude (in units of σ_a)					
			ARL(0)	ARL(1)	ARL(2)	ARL(3)	ARL(4)	ARL(5)
3.9	0.0127	2.0805	500	31.6	14.1	8.90	6.40	4.96
	(0.0229)	(2.2216)	(500)	(29.3)	(12.7)	(8.15)	(6.04)	(4.86)
3.7	0.0390	2.5289	500	**27.6**	10.9	6.55	4.55	3.43
	(0.0716)	(2.5898)	(500)	(**28.1**)	(9.98)	(5.98)	(4.33)	(3.44)
3.5	0.0667	2.6977	500	27.8	9.80	5.56	3.75	2.77
	(0.1243)	(2.7160)	(500)	(30.7)	(9.17)	(5.12)	(3.61)	(2.83)
3.3	0.0959	2.8020	500	30.0	9.28	5.07	3.29	2.39
	(0.1814)	(2.7979)	(500)	(35.0)	(**8.97**)	(4.66)	(3.19)	(2.47)
3.1	0.1268	2.8685	500	32.3	**9.05**	4.69	2.99	2.10
	(0.2432)	(2.8593)	(500)	(39.9)	(9.01)	(4.41)	(2.90)	(2.25)
2.9	0.1594	2.9188	500	35.4	9.18	4.44	2.74	1.84
	(0.3099)	(2.9044)	(500)	(43.8)	(9.41)	(4.22)	(2.69)	(2.08)
2.7	0.1940	2.9571	500	40.6	9.37	4.32	2.52	1.63
	(0.3819)	(2.9430)	(500)	(51.0)	(10.0)	(4.15)	(2.53)	(1.90)
2.5	0.2308	2.9848	500	44.5	9.62	4.17	2.34	1.46
	(0.4593)	(2.9720)	(500)	(55.0)	(10.3)	(4.07)	(2.39)	(1.70)
2.3	0.2698	3.0064	500	48.5	10.1	**4.09**	**2.16**	**1.34**
	(0.5420)	(2.9950)	(500)	(59.1)	(10.9)	(**4.04**)	(**2.25**)	(**1.54**)

Table 3 Comparison of ARL performance for EWMAs on e_t and on x_t with common sensitivity to ϕ and in-control ARL of 500. x_t is ARMA(1,1) with $\phi = 0.87$ and $\theta = 0.48$. The values for the EWMA on x_t are in parentheses.

$S(\phi)$	$\mid S(\theta) \mid$	λ	L	ARL(0)	ARL(1)	ARL(2)	ARL(3)	ARL(4)	ARL(5)
						mean shift magnitude (in units of σ_a)			
14.0	3.75 (3.81)	0.0127 (0.0263)	2.0805 (2.0791)	500 (500)	**69.7** (**70.1**)	26.6 (25.4)	14.6 (14.7)	9.08 (10.4)	6.03 (8.11)
13.0	3.68 (3.77)	0.0233 (0.0514)	2.3389 (2.2777)	500 (500)	70.0 (74.2)	23.8 (22.9)	12.2 (12.2)	7.19 (8.25)	4.57 (6.32)
12.0	3.59 (3.72)	0.0354 (0.0842)	2.4954 (2.4086)	500 (500)	73.4 (80.9)	22.8 (22.3)	10.8 (10.8)	5.87 (6.87)	3.70 (5.14)
11.0	3.50 (3.65)	0.0493 (0.1280)	2.6101 (2.5152)	500 (500)	78.3 (87.5)	**22.0** (**21.9**)	9.55 (9.59)	4.96 (5.80)	3.09 (4.25)
10.0	3.39 (3.56)	0.0654 (0.1883)	2.6921 (2.6133)	500 (500)	85.6 (95.3)	22.5 (22.8)	8.83 (8.90)	4.29 (5.01)	2.68 (3.56)
9.0	3.27 (3.42)	0.0844 (0.2728)	2.7653 (2.7068)	500 (500)	93.5 (100)	22.7 (23.1)	8.35 (8.44)	3.81 (4.37)	2.38 (2.92)
8.0	3.13 (3.22)	0.1071 (0.3906)	2.8297 (2.8038)	500 (500)	105 (108)	24.1 (24.0)	7.97 (8.09)	3.47 (3.81)	2.15 (2.44)
7.0	2.96 (2.94)	0.1347 (0.5472)	2.8823 (2.8911)	500 (500)	117 (114)	26.1 (25.1)	7.84 (**7.94**)	3.10 (3.39)	1.94 (2.00)
6.0	2.76 (2.56)	0.1690 (0.7367)	2.9302 (2.9646)	500 (500)	131 (119)	28.9 (26.6)	**7.68** (8.00)	2.81 (**3.07**)	1.70 (1.63)
5.0	2.53 (2.11)	0.2126 (0.9432)	2.9723 (3.0204)	500 (500)	149 (129)	32.7 (28.2)	7.94 (8.39)	2.63 (2.93)	1.51 (**1.44**)
4.0	2.25	0.2701	3.0065	500	172	38.9	8.55	**2.36**	**1.32**

Table 4 Comparison of ARL performance for EWMAs on e_t and on x_t with common sensitivity and in-control ARL of 500. x_t is ARMA(1,1) with $\phi = 0.7$ and $\theta = 0.3$. The values for the EWMA on x_t are in parentheses.

| $S(\phi)$ | $\lvert S(\theta) \rvert$ | λ | L | mean shift magnitude (in units of σ_a) | | | | | |
				ARL(0)	ARL(1)	ARL(2)	ARL(3)	ARL(4)	ARL(5)
6.4	2.81	0.0123	2.0665	500	37.5	16.1	9.79	6.78	5.06
	2.83	(0.0227)	(2.1649)	(500)	(35.2)	(14.7)	(9.25)	(6.83)	(5.44)
6.0	2.73	0.0323	2.4636	500	**34.0**	12.9	7.39	4.87	3.53
	2.77	(0.0610)	(2.4823)	(500)	(**34.5**)	(11.9)	(7.01)	(5.01)	(3.97)
5.6	2.64	0.0541	2.6376	500	34.5	11.5	6.21	3.99	2.82
	2.71	(0.1052)	(2.6240)	(500)	(37.4)	(10.9)	(5.97)	(4.16)	(3.26)
5.2	2.55	0.0780	2.7427	500	36.4	11.0	5.55	3.44	2.40
	2.64	(0.1565)	(2.7157)	(500)	(41.4)	(**10.6**)	(5.35)	(3.62)	(2.79)
4.8	2.45	0.1045	2.8241	500	40.1	**10.7**	5.06	3.03	2.15
	2.55	(0.2163)	(2.7915)	(500)	(47.0)	(10.7)	(4.91)	(3.19)	(2.43)
4.4	2.34	0.1339	2.8810	500	44.7	10.8	4.76	2.76	1.92
	2.45	(0.2859)	(2.8471)	(500)	(52.1)	(10.9)	(4.64)	(2.89)	(2.21)
4.0	2.22	0.1667	2.9275	500	50.1	11.0	4.51	2.52	1.70
	2.33	(0.3666)	(2.9050)	(500)	(57.5)	(11.3)	(4.39)	(2.66)	(1.98)
3.6	2.09	0.2035	2.9656	500	56.5	11.4	4.34	2.34	1.51
	2.18	(0.4596)	(2.9513)	(500)	(62.3)	(11.7)	(4.27)	(2.42)	(1.72)
3.2	1.95	0.2453	2.9941	500	63.9	12.4	4.30	2.16	1.38
	2.01	(0.5652)	(2.9863)	(500)	(67.1)	(12.8)	(4.33)	(2.23)	(1.50)
2.8	1.80	0.2929	3.0165	500	72.9	13.4	**4.25**	1.97	1.27
	1.80	(0.6830)	(2.9050)	(500)	(73.2)	(13.2)	(**4.22**)	(2.03)	(1.32)
2.4	1.62	0.3478	3.0377	500	83.9	15.2	4.30	1.83	1.19
	1.57	(0.8112)	(2.9513)	(500)	(79.2)	(14.3)	(4.26)	(1.89)	(1.22)
2.0	1.43	0.4118	3.0569	500	96.9	17.3	4.49	**1.75**	**1.14**
	1.30	(0.9474)	(2.9863)	(500)	(85.8)	(15.3)	(4.48)	(**1.83**)	(**1.16**)

II.7 Chapter Summary

In this chapter, we have developed analytical results for the sensitivity of EWMA control charts on autocorrelated data and on the residuals of an ARMA model of the process. For an EWMA on x_t, or more generally a linear filter on x_t, the sensitivities are expressed in terms of the nominal autocorrelation function of the filter output. For the residual-based EWMA, the sensitivities reduce to relatively simple expressions of the nominal ARMA polynomials $\hat{\Phi}(B)$ and $\hat{\Theta}(B)$ and the EWMA parameter.

The analytical results and the simulation results both indicate that although the residual-based EWMA is sensitive to modeling errors, it is generally less sensitive than the EWMA on x_t. Likewise, it is no more sensitive than a control chart applied to the closed-loop process output after attempting to remove the autocorrelation with feedback control.

CHAPTER III

ROBUST DESIGN OF RESIDUAL-BASED CONTROL CHARTS FOR AUTOCORRELATED PROCESSES: WORST CASE APPROACH*

III.1 Introduction

In Chapter II, it was shown that both methods for control charts — residual-based EWMA control charts and EWMA control charts directly on x_t — were sensitive to the ARMA modeling errors. Moreover, applying the EWMA control charts directly to x_t could not be an effective option to resolve the sensitivity problem of the residual-based EWMA control charts. Rather, the residual-based EWMA chart was more robust with respect to the ARMA modeling errors than the alternative EWMA chart for the same value of λ. Consequently, a reasonable strategy can be to use a residual-based EWMA, but widen the control limits to some extent, in order to account for the model uncertainty when we apply to control charts in practice. In this chapter, we develop a robust design method for the residual-based EWMA control chart in the presence of ARMA modeling errors.

It is assumed that the autocorrelated process data, x_t, follows (1.1) and the in-control process mean has been subtracted so that x_t is 0-mean until there is a shift. For notational convenience the results in this chapter are derived for ARMA processes, although a straightforward extension to autoregressive integrated moving average (ARIMA) processes is discussed in Chapter III.2.

When ARMA modeling errors exist, the residuals, e_t, generated via the estimated model behave as the ARMA($p+q$, $p+q$) process

$$e_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} x_t = \frac{\hat{\Phi}(B)\Theta(B)}{\hat{\Theta}(B)\Phi(B)} a_t .$$ (3.1)

and are no longer iid. When the EWMA statistic is applied to the residuals, then z_t can be written as the ARMA($p+q+1$, $p+q$) proce

$$z_t = \frac{1-v}{1-vB} e_t = \frac{(1-v)\hat{\Phi}(B)\Theta(B)}{(1-vB)\hat{\Theta}(B)\Phi(B)} a_t$$ (3.2)

From Chapter II, the standard residual-based EWMA chart design is to set the upper control limit (UCL) and lower control limit (LCL) on z_t at

$$\{\text{LCL,UCL}\} = \pm L \hat{\sigma}_z ,$$ (3.3)

where $\hat{\sigma}_z = \hat{\sigma}_a (1-v)^{1/2}(1+v)^{-1/2}$ is the steady-state standard deviation of z_t assuming the estimated model is perfect. To improve the sensitivity to mean shifts that occur when the control chart is first initiated, time-varying control limits that gradually widen to the steady-state limits (3.3) can also be used (Montgomery, 2001). This study considers only constant steady-state control limits.

Let σ_z^2 denote the actual variance of the EWMA statistic (3.2), which is a function of the true parameters and their estimates. As discussed in Chapter I.1, if the true and estimated models are such that the residual autocorrelation is positive, then σ_z^2 will be larger (possibly much larger) than believed, the control chart will be frequently interrupted by false alarms.

To illustrate the effects of modeling errors, suppose that x_t is an AR(1) process with $\phi_1 = 0.9$ and $\sigma_a^2 = 1.0$ and that the estimated parameters are $\hat{\phi}_1 = 0.85$ and $\hat{\sigma}_a^2 = 1.0$. Using an EWMA with $\lambda = 0.1$ and treating the estimates as perfect, the assumed EWMA variance is $\sigma_z^2 = \hat{\sigma}_a^2 (1-v)(1+v)^{-1} = 0.053$. For a desired in-control ARL of 500, $L = 2.814$ (Lucas and Saccucci, 1990) and the control limits $\pm L \hat{\sigma}_z = \pm 0.647$ would be used. Using (2.2) for calculating the variance of z_t, however, it can be shown that the actual EWMA variance is $\sigma_z^2 = 0.084$ — roughly 60% larger than the assumed variance. If the control limits based on the assumed variance are used, then Monte Carlo simulation (refer to Chapter III.4 for details) reveals the actual in-control ARL is approximately 165, which is substantially shorter than intended. Figure 4, which shows the EWMA statistic for 500 simulated observations with the ± 0.647 control limits, illustrates the frequent false alarms that result in this situation.

To protect against a situation in which the in-control ARL is considerably shorter than desired, a logical precaution is to use control limits that are wider than those used when the model is assumed perfect. This chapter presents a method for widening the

EWMA control limits based on the following "worst-case" design approach. For a specified λ and a given set of ARMA parameter estimates, (3.2) implies σ_z is a function of the true, unknown parameters. Considering the uncertainty in the true parameters, Chapter III.2 derives an approximate upper one-sided $1-\alpha$ confidence interval for σ_z for some user-selected $0<\alpha<1$. Let $\sigma_{z,\alpha}$ denote the upper boundary of this confidence interval, which can be viewed as a worst-case (maximum) value for the true EWMA standard deviation. $\sigma_{z,\alpha}$ will be represented in the form that involves the sensitivity results of (2.12) and (2.13) for the residual-based EWMA derived in Chapter II.
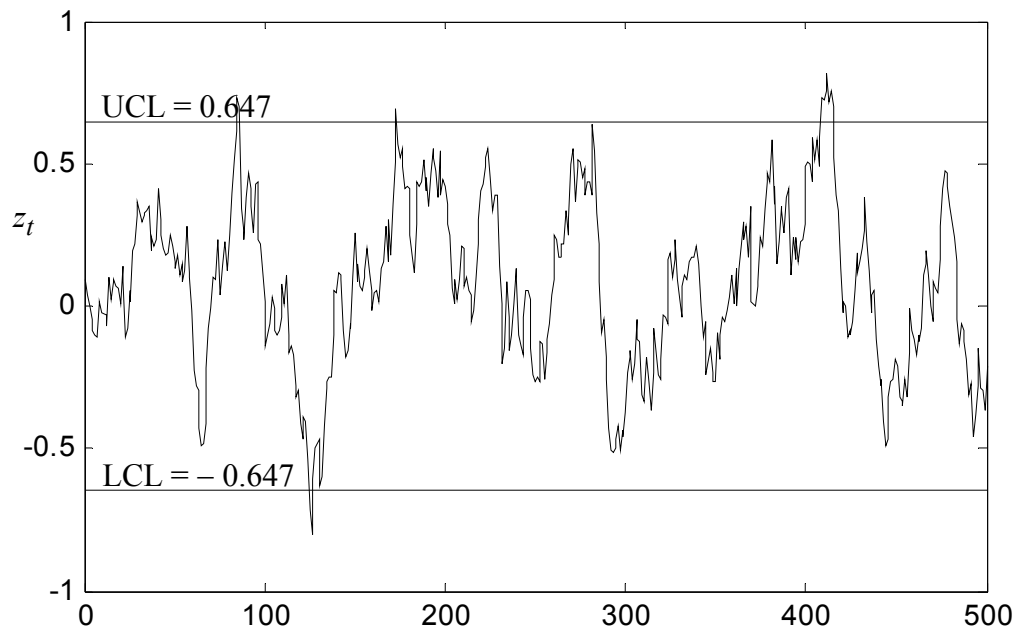
Figure 4 Example EWMA chart for an in-control AR(1) process with ϕ_1 underestimated. The desired in-control ARL is 500, whereas the actual ARL is much lower due to frequent false alarms.

The proposed method is to monitor the EWMA statistic (3.2), but to use the worst-case control limits

$$\{LCL, UCL\} = \pm L\sigma_{z,\alpha} \tag{3.4}$$

instead of the standard control limits (3.3). Chapter III.3 discusses guidelines for selecting the design parameters L, λ, and α. L can be chosen so that the worst-case ARL (roughly, the in-control ARL that would result if σ_z assumed its worst-case value) approximately equals some desired ARL value specified by the user. Widened control limits will inevitably increase the out-of-control ARL for any size mean shift and reduce the power of the chart. Chapter III.4 discusses this drawback of the worst-case design approach and illustrate with examples. It also discusses sample size requirements and compare the EWMA with a Shewhart individual chart, which is less powerful for small to moderate mean shifts but more robust to modeling errors.

III.2 Worst Case EWMA Variance

Form Chapter II.2, the EWMA statistic (3.2) can be rewritten as (2.1) where $G(B)$ $= (1-v)(1-vB)^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)\Phi^{-1}(B)\Theta(B) = \sum_{j=0}^{\infty} g_j B^j$ and $\{g_j: j = 0, 1, 2, \ldots\}$ are the impulse response coefficients of the ARMA($p+q+1,p+q$) transfer function $G(B)$. For a fixed set of ARMA parameters and their estimates, the EWMA variance is calculated by (2.2).

Define the ARMA parameter vector $\gamma = [\phi_1 \ \phi_2 \ldots \phi_p \ \theta_1 \ \theta_2 \ldots \theta_q \ \sigma_a^2]^T$, and let $\hat{\gamma}$ denote a point estimate. Note that we also consider the modeling error of σ_a^2 in this

chapter, therefore σ_a^2 is additionally included in γ. To find an approximate confidence

interval for σ_z, we use a first-order Taylor approximation of the ratio $\sigma_z^2/\hat{\sigma}_z^2$ about $\hat{\gamma}$

$= \gamma$. If the parameter error vector is defined as $\tilde{\gamma} = \hat{\gamma} - \gamma$, then the first-order Taylor

approximation is

$$\sigma_z^2/\hat{\sigma}_z^2 \cong 1 + V^T \tilde{\gamma} , \qquad (3.5)$$

where

$$V = \left[\frac{-2v}{\Phi(v)} \quad \frac{-2v^2}{\Phi(v)} \quad \cdots \quad \frac{-2v^p}{\Phi(v)} \quad \frac{2v}{\Theta(v)} \quad \frac{2v^2}{\Theta(v)} \quad \cdots \quad \frac{2v^q}{\Theta(v)} \quad -\sigma_a^{-2} \right]^T$$

with $\Phi(v) = \Phi(B)|_{B=v} = 1 - \phi_1 v - \phi_2 v^2 - \ldots - \phi_p v^p$, and $\Theta(v) = \Theta(B)|_{B=v} =$

$1 - \theta_1 v - \theta_2 v^2 - \ldots - \theta_q v^q$. When (2.13) and (2.14) are compared to V (except for the

last element), we can see that signs become reversed and the estimates of ARMA

polynomials are replaced by true polynomials in V. This is because we use the Taylor

approximation about $\hat{\gamma} = \gamma$ in this chapter instead of $\gamma = \hat{\gamma}$. In Chapter II, we viewed γ

as a random variable and differentiated with respect to γ to derive the sensitivity

measure. On the contrary, if we differentiate $G(B)$ with respect to $\hat{\gamma}$, the final result

changes to the shown form because the impulse response coefficient, g_j, becomes the

linear function of $\hat{\Phi}(B)$ from $G(B) = (1-v)(1-vB)^{-1}\hat{\Theta}^{-1}(B)\hat{\Phi}(B)\Phi^{-1}(B)\Theta(B)$. This

is discussed more specifically in Chapter IV.5.2. For σ_a^2, since the ratio, $\sigma_z^2/\hat{\sigma}_z^2$, is

$(1-v)^{-1}(1+v)\hat{\sigma}_a^{-2}\sigma_a^2\sum_{j=0}^{\infty}g_j^2$, differentiating the ratio with respect to $\hat{\sigma}_a^2$, and

evaluating the result at σ_a^2 gives $-\left(1-v\right)^{-1}\left(1+v\right)\sigma_a^{-2}\sum_{j=0}^{\infty}g_j^2\Big|_{\hat{\sigma}_a^2=\sigma_a^2}$. Therefore, the

result is $-\sigma_a^{-2}$ since $\sum_{j=0}^{\infty}g_j^2\Big|_{\hat{\sigma}_a^2=\sigma_a^2}=\left(1-v\right)\left(1+v\right)^{-1}$.

Let N denote the number of observations in the sample used to estimate the ARMA parameters. For most estimation methods, the distribution of $\tilde{\gamma}$ for large N is approximately multivariate normal with mean 0 and some covariance matrix Σ_γ that is inversely proportional to N (Box, et al., 1994; Brockwell and Davis, 1991). Commercial statistical software packages for ARMA modeling often provide an estimate $\hat{\Sigma}_\gamma$ of the covariance along with the parameter estimates. Alternatively, the method outlined in Appendix A may be used to calculate $\hat{\Sigma}_\gamma$ when only the parameter estimates are available. Closed-form expressions for $\hat{\Sigma}_\gamma$ are also provided in Appendix A for the special case of first-order ARMA processes.

Using the multivariate normal approximation to the distribution of $\tilde{\gamma}$, the ratio $\sigma_z^2/\hat{\sigma}_z^2$ in (3.5) is approximately normally distributed with mean 1 and variance $V^T\Sigma_\gamma V$. Thus, for any probability $0 < \alpha < 1$,

$$1-\alpha \cong Pr\left[\sigma_z^2/\hat{\sigma}_z^2 \le 1 + z_\alpha(V^T\Sigma_\gamma V)^{1/2}\right] = Pr\left[\sigma_z \le \hat{\sigma}_z\left\{1 + z_\alpha(V^T\Sigma_\gamma V)^{1/2}\right\}^{1/2}\right],$$

where z_α denotes the upper α percentile of the standard normal distribution. Substituting $\hat{\Sigma}_\gamma$ and

$$\hat{V} = \left[\frac{-2v}{\hat{\Phi}(v)} \quad \frac{-2v^2}{\hat{\Phi}(v)} \quad \cdots \quad \frac{-2v^p}{\hat{\Phi}(v)} \quad \frac{2v}{\hat{\Theta}(v)} \quad \frac{2v^2}{\hat{\Theta}(v)} \quad \cdots \quad \frac{2v^q}{\hat{\Theta}(v)} \quad -\hat{\sigma}_a^{-2} \right]^T \qquad (3.6)$$

for Σ_γ and V leads to the approximate $1-\alpha$ confidence interval

$$\sigma_z \leq \sigma_{z,\alpha} = \hat{\sigma}_z \left\{ 1 + z_\alpha (\hat{V}^T \hat{\Sigma}_\gamma \hat{V})^{1/2} \right\}^{1/2} \qquad (3.7)$$

for the EWMA standard deviation. After selecting L as described in the following chapter, $\sigma_{z,\alpha}$ can be used in the worst-case control limits (3.4).

The Taylor approximation (3.5) has an interesting interpretation when the process is ARMA(1,1). In this case, (3.5) reduces to

$$\sigma_z^2 \cong \hat{\sigma}_z^2 \left\{ 1 - \frac{2v(\hat{\phi}_1 - \phi_1)}{1 - \phi_1 v} + \frac{2v(\hat{\theta}_1 - \theta_1)}{1 - \theta_1 v} - \frac{\hat{\sigma}_a^2 - \sigma_a^2}{\sigma_a^2} \right\}.$$

The EWMA variance increases (relative to the assumed value $\hat{\sigma}_z^2$) when ϕ_1 is underestimated ($\hat{\phi}_1 < \phi_1$) and/or θ_1 is overestimated ($\hat{\theta}_1 > \theta_1$). The reason is that the autocorrelation of x_t is underestimated in this situation, resulting in residuals with positive autocorrelation. When the residuals are positively autocorrelated, the variance of their EWMA is larger than if the residuals were iid. This was discussed in more detail in Adams and Tseng (1998). The foregoing equation also indicates that the effects of parameter estimation errors are larger for larger values of v. In the limiting case with $v = 0$ (a Shewhart individual chart on the residuals), errors in estimating ϕ_1 and θ_1 have very little effect on the EWMA variance, which is further discussed in Chapter III.4.3.

The confidence interval (3.7) and the expressions for $\hat{\Sigma}_\gamma$ in Appendix A are also valid for ARIMA(p,1,q) processes of the form $x_t = (1-B)^{-1}\Phi^{-1}(B)\Theta(B)a_t$. The reason is that when estimating the parameters of an ARIMA model, one fits an ARMA model to the differenced data $(1-B)x_t$. Since the residuals are still generated via (3.1) with x_t replaced by the differenced data, the EWMA statistic follows the same ARMA($p+q+1$,$p+q$) model (3.2). The parameter errors therefore have the exact same effect on the EWMA variance as in the ARMA case.

III.3 Selecting Design Parameters

When designing an EWMA chart for iid data with no consideration of model uncertainty, the parameters λ and L are often jointly selected to minimize the out-of-control ARL for a specified mean shift, while ensuring the in-control ARL equals some desired value. Lucas and Saccucci (1990) provide tables for selecting values of λ and L that are optimal in this sense. For a residual-based EWMA with autocorrelated data, optimally selecting λ and L is complicated even when perfect models are assumed. The optimal λ and L depend on many factors, including the desired in-control ARL, the specified mean shift of interest, and the ARMA parameters. For first-order autoregressive models, Lu and Reynolds (1999) provide tables for selecting the optimal λ and L for the specific cases of $\phi_1 = 0.4$ and $\phi_1 = 0.8$ with a desired in-control ARL of 370. When considering model uncertainty as in this chapter, jointly selecting λ and L to satisfy some optimality criterion is prohibitively complex.

In light of this, it is recommended that one first select λ as if the estimated model were perfect. The rule-of-thumb $0.05 < \lambda \leq 0.5$ (Lu and Reynolds, 1999) may be used, where it is understood that smaller λ values result in better detection of small mean shifts, but slower detection of large shifts. For more detailed guidelines, the reader is referred to the thorough discussions in Lucas and Saccucci (1990) and Lu and Reynolds (1999).

After specifying λ, suppose that the tables of Lucas and Saccucci (1990) are used to select L based on some desired in-control ARL (denoted ARL_d). If used in the standard EWMA control limits (3.3), this value of L would provide the desired ARL when there is no model uncertainty and the residuals are iid. With model uncertainty considered, using the same value of L in the worst-case EWMA control limits (3.4) is recommended. If the EWMA standard deviation σ_z happens to coincides with its worst-case value $\sigma_{z,\alpha}$, then the control limits (3.4) will provide an in-control ARL that approximately equals the desired value ARL_d. The examples in Chapter III.4 indicate that this choice of L also results in an appealing Bayesian interpretation of the control chart: If an appropriate posterior distribution for the ARMA parameters is considered, then the posterior probability that the ARL is less than ARL_d is reasonably close to the α value specified in the confidence interval on σ_z.

Using a slightly smaller value of L in the control limits (3.4) also might have been considered for the following reason. When there are no modeling errors, and the standard control limits (3.3) are used, the value of L that provides a desired in-control

ARL depends on λ. This is primarily because the autocorrelation of the EWMA statistic z_t depends on λ. As λ decreases, the autocorrelation of z_t increases, and the in-control ARL increases for any fixed L. Consequently, as λ decreases, smaller values of L will provide the same in-control ARL. When modeling errors are present, the errors also affect the autocorrelation of z_t. When the true parameters are such that σ_z coincides with $\sigma_{z,\alpha}$, the autocorrelation of the residuals will generally be positive, and the autocorrelation of z_t will be larger than when there are no modeling errors. Consequently, a slightly smaller value of L may provide the desired ARL when σ_z coincides with $\sigma_{z,\alpha}$. On the other hand, a first-order Taylor approximation of the EWMA variance was also used in developing the expression for $\sigma_{z,\alpha}$. This approximation tends to underestimate the EWMA variance, and the resulting $\sigma_{z,\alpha}$ is slightly smaller than what would result from a more exact confidence interval. Since the control limits (3.4) are the product of L and $\sigma_{z,\alpha}$, the effects of the Taylor approximation are partially compensated by taking L directly from the tables Lucas and Saccucci (1990) as recommended, as opposed to using a slightly smaller value.

Note that the ARL_d that one specifies in the design procedure should be viewed as a worst-case ARL that results when the EWMA variance equals its worst-case value (within the $1-\alpha$ confidence interval). If the true ARMA parameters and the EWMA variance are close to their estimates, the ARL will generally be larger than ARL_d. To avoid overly conservative control limits, this should be kept in mind when selecting

the remaining design parameter α. A small value such as $\alpha = 0.01$ may widen the

control limits to an extent that makes it difficult to detect most mean shifts of interest.

This tradeoff in using the worst-case control limits is discussed in more detail in

Chapter III.4.1 and III.4.2, with a recommended range $0.1 \le \alpha \le 0.3$.

The design procedure is illustrated with the Series A data from Box, et al. (1994),

which are $N = 197$ concentration measurements from a chemical process. Box, et al.

(1994) found that an ARMA(1,1) model fit the data well, and the estimated parameters

were (omitting their subscripts) $\hat{\phi} = 0.87$, $\hat{\theta} = 0.48$, and $\hat{\sigma}_a^2 = 0.098$. Using Equation

(A.4), the estimated parameter covariance is

$$\hat{\Sigma}_\gamma = \begin{bmatrix} 2.75 & 3.64 & 0 \\ 3.64 & 8.71 & 0 \\ 0 & 0 & 0.098 \end{bmatrix} \times 10^{-3}.$$

If $\lambda = 0.1$ and $ARL_d = 500$ are selected, the tables of Lucas and Saccucci (1990)

indicate that $L = 2.814$ should be used. Since $\hat{\sigma}_z = \hat{\sigma}_a (1-\nu)^{1/2}(1+\nu)^{-1/2} = 0.0718$, the

standard control limits (3.3) become $\pm L \hat{\sigma}_z = \pm 0.202$. If $\alpha = 0.1$ is also selected, then

(3.6) and (3.7) result in $\hat{V} = [-8.29 \quad 3.17 \quad -10.20]^T$, and $\sigma_{z,\alpha} = 0.0849$. The worst-

case control limits (3.4) are therefore $\pm L\sigma_{z,\alpha} = \pm 0.239$, which are 18% wider than the

standard control limits.

Figure 5 shows an EWMA control chart applied to 500 simulated observations

from the process when the true parameters assume the values $\phi = 0.917$, $\theta = 0.491$,

and $\sigma_a^2 = 0.102$. These parameter values were chosen because the resulting Taylor

approximation (3.5) of σ_z^2 (with V replaced by $\hat{V}$) equals the worst-case value $\sigma_{z,\alpha}^2$.

One can also show that of all parameter combinations that result in a Taylor approximation equal to $\sigma_{z,\alpha}^2$, these values have the highest likelihood (minimize $\tilde{\gamma}^T \hat{\Sigma}_\gamma^{-1} \tilde{\gamma}$). Both the standard and the worst-case control limits are shown in Figure 5. Since the mean of x_t was held at 0 throughout the simulation, all cases where the EWMA statistic fell outside the control limits were false alarms. The standard control limits resulted in false alarms around timesteps 50, 275, and 425, whereas the worst-case control limits eliminated the first two of these. Monte Carlo simulation is used in the following chapter to provide a more comprehensive analysis of the control chart performance.
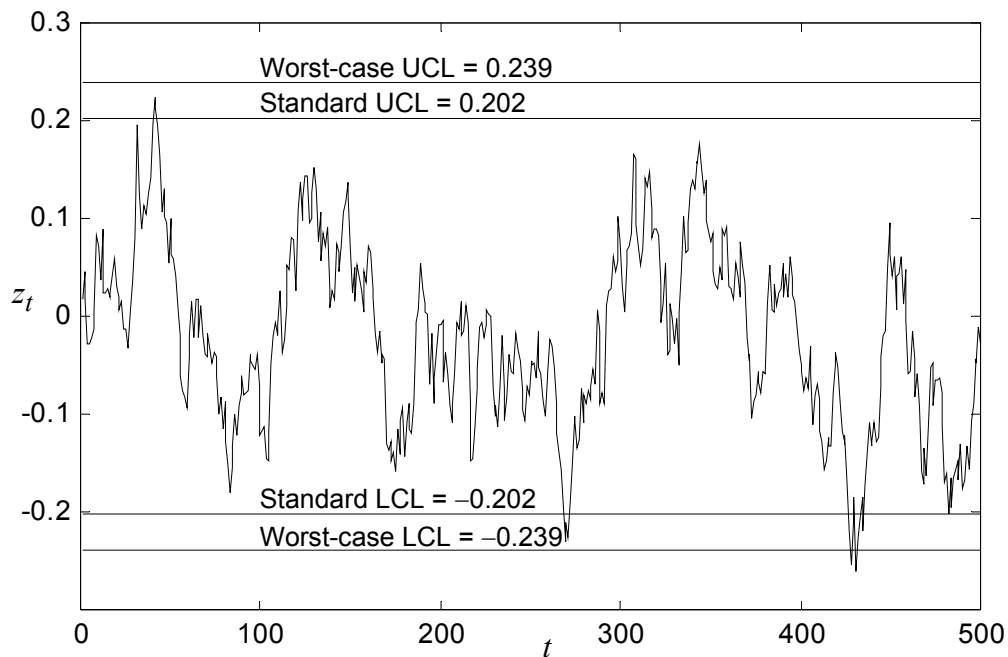
Figure 5 Example EWMA chart with standard and worst-case control limits, when σ_z coincides with its worst-case value $\sigma_{z,\alpha}$.
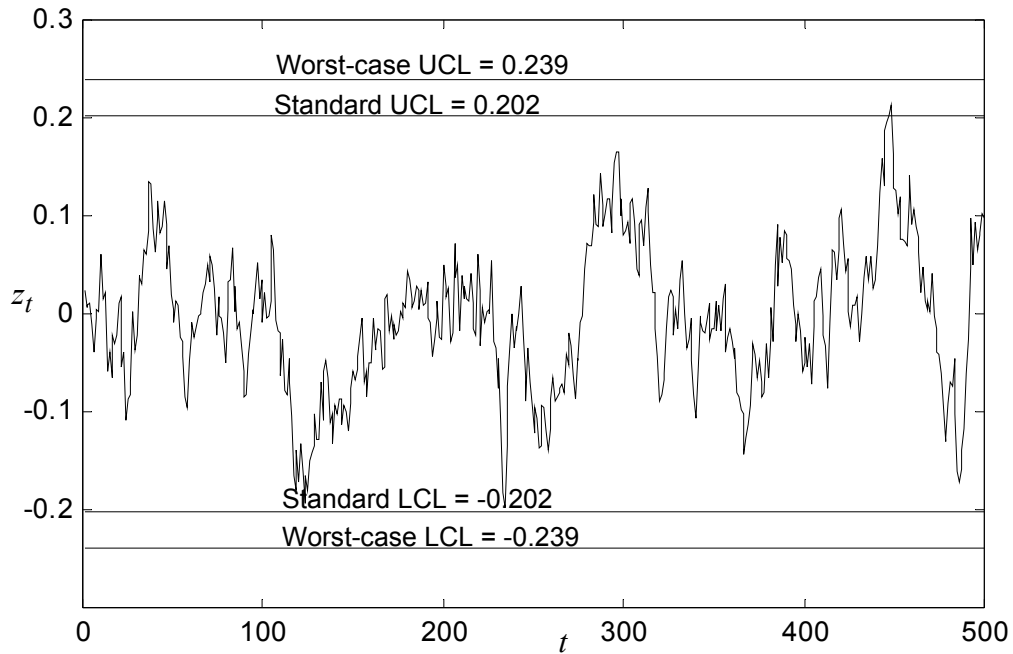
Figure 6 Example EWMA chart with standard and worst-case control limits, when σ_z coincides with $\hat{\sigma}_z$.

Figure 6, which is similar to Figure 5 except that the true ARMA parameters were chosen to coincide with their estimates, illustrates one drawback of using the worst-case control limits: If the true parameters happen to fall sufficiently close to their estimates, then the standard control limits provide the desired in-control ARL. The worst-case control limits are unnecessarily wide in this case, which inevitably decreases the power of the control chart. This is an inherent consequence of the worst-case design approach, which is intended to guard against the situation where the true parameters are not "sufficiently" close to their estimates. To mitigate this drawback, using both sets of control limits for the EWMA chart is recommended. An observation falling outside the worst-case control limits provides strong evidence that the process

has changed. An observation falling within the worst-case control limits but outside the standard limits should be interpreted with more caution; it could mean that either the process has changed or that the ARMA parameters differ from their estimates. Chapter III.4 provides a detailed discussion of the tradeoffs involved in the worst-case design approach.

III.4 Discussions

Monte Carlo simulation is used with exactly the same manner, which was explained in Chapter II.6, throughout this chapter to investigate the ARL performance of EWMA charts with standard and worst-case control limits when the parameters differ from their estimates. Only difference is that modeling errors are considered in this chapter. The EWMA for the residuals was calculated via (3.2), with z_0 initialized at 0. A signal occurred when z_t fell outside the control limits.

III.4.1 Bayesian Interpretations

Consider a Bayesian alternative to the worst-case design approach, where some posterior distribution for γ is assumed (given the data from which the parameters are estimated) and the control limits are selected to provide a desired *average* ARL with respect to the posterior distribution of γ. This chapter discusses why designing the control chart based on an average ARL would actually lead to control limits that are narrower than the standard limits. In addition, a Bayesian analysis is considered to investigate the posterior probability that the ARL is less than ARL_d when the worst-

case control limits are used. For the examples considered here, this probability is reasonably close to the value of α specified in the confidence interval. For analysis purposes, it is assumed the posterior distribution of γ is approximately multivariate normal with mean $\hat{\gamma}$ and covariance $\hat{\Sigma}_\gamma$ (see Appendix A). This can be viewed as an asymptotic approximation when the prior distribution of γ is noninformative.

Reconsider the ARMA(1,1) example introduced in Chapter III.3, where the estimated parameters were $\hat{\phi} = 0.87$, $\hat{\theta} = 0.48$, and $\hat{\sigma}_a^2 = 0.098$. For simplicity, uncertainty in σ_a^2 is neglected by modifying the earlier expression for $\hat{\Sigma}_\gamma$ so that its lower-right element (i.e., the variance of $\hat{\sigma}_a^2$) is 0. This results in $\sigma_{z,\alpha} = 0.0842$ and worst-case control limits $\pm L\sigma_{z,\alpha} = \pm 0.237$, which are only slightly narrower than when we also considered uncertainty in σ_a^2. Figure 7 shows contour plots of the ARL as a function of ϕ and θ for $\sigma_a^2 = \hat{\sigma}_a^2$. Panel (a) is the in-control ARL contours for the standard EWMA with control limits ± 0.202. The parameter estimates are indicated by the * symbol. Since the EWMA was designed with $ARL_d = 500$, the ARL = 500 contour passes through the parameter estimates. Numerical integration of the ARL with respect to the assumed posterior density of γ gives a rough approximation of 730 for the average ARL of the EWMA chart with standard control limits. Somewhat surprisingly, this is larger than the desired ARL of 500 that results when the model is perfect. It may be concluded that an average ARL of 500 could be achieved with control limits that are even narrower than the standard control limits.
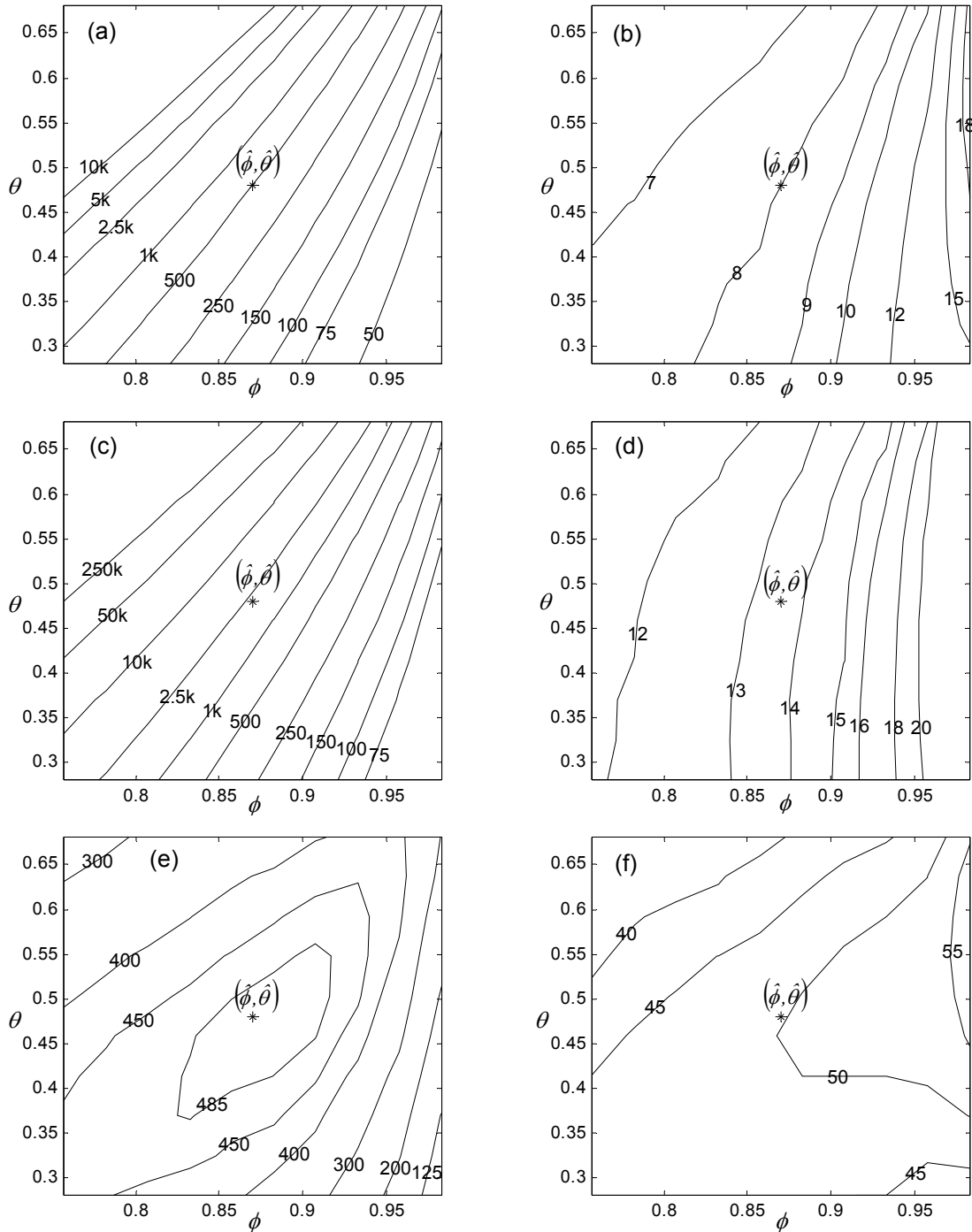
Figure 7 ARL contours as a function of ϕ and θ for the ARMA(1,1) example. Panels (a), (c), and (e) show the in-control ARLs for the standard EWMA, worst-case EWMA, and Shewhart chart, respectively. Panels (b), (d), and (f) show the out-of-control ARLs for the three charts when the mean shift magnitude is $3\sigma_a$.

The reason the average ARL is larger than 500 is that the ARL is a highly skewed function of ϕ and θ, as can be seen in Figure 7(a). For $\phi < \hat{\phi}$ and $\theta > \hat{\theta}$, the ARL increases dramatically. The average ARL is misleading, however, since the ARL may decrease to unacceptably small values for $\phi > \hat{\phi}$ and $\theta < \hat{\theta}$. Numerical integration (of the posterior density over the ARL < 250 region) also reveals there is a 0.24 probability the ARL is less than 250, which is only half the desired ARL. Likewise, there is a 0.11 probability the ARL is less than 150.

Figure 7(c) shows analogous in-control ARL contours for the EWMA chart with worst-case control limits ±0.237. With the worst-case control limits, the probability the ARL is less than the desired value 500 is approximately 0.13, which is reasonably close to the $\alpha = 0.1$ value selected when the chart was designed. Moreover, the probability the ARL is less than 250 is only 0.05, compared to the 0.24 probability with the standard control limits. The worst-case control limits clearly provide adequate protection against an unacceptably short in-control ARL. An additional benefit is that when the parameters coincide with their estimates, the in-control ARL will be even larger than the desired value. From Figure 7(c), the in-control ARL in this case is roughly 2000, compared to an ARL of 500 with the standard control limits. The obvious disadvantage of widening the control limits, which is discussed in the following chapter, is the resulting decrease in the power of the chart for detecting mean shifts.

III.4.2 In-Control versus Out-of-Control ARL Trade-off

For the same ARMA(1,1) example introduced in Chapter III.3 and continued in Chapter III.4.1, Figures 7(b) and 7(d) show the out-of-control ARL contours with a mean shift of magnitude $3\sigma_a$. Figure 7(b) is for the standard control limits $\pm$ 0.202, and Figure 7(d) is for the worst-case control limits $\pm$ 0.237. The worst-case control limits increase the out-of-control ARL by approximately 60% for most combinations of ϕ and θ. Note that even with the standard control limits, the ARL is approximately 8.0 when the ARMA parameters equal their estimates, which may seem large for a mean shift of $3\sigma_a$. After the initial occurrence of the mean shift, however, the mean of the residuals rapidly approaches a steady-state value of only $0.75\sigma_a$. Superville and Adams (1994) and Apley and Shi (1999) discussed this "forecast recovery" phenomenon in detail. Table 5 presents the out-of-control ARL values for other mean shifts for the specific case that the ARMA parameters coincide with their estimates. It also provides results for the Shewhart individual chart, discussed in Chapter III.4.3. Widening the control limits clearly has a negative impact on the out-of-control ARL, particularly for small mean shifts. For a mean shift of size σ_a, which results in a steady-state residual mean of only $0.25\sigma_a$, widening the control limits causes the out-of-control ARL to increase from 101 to 247. This is understandable, given that the in-control ARL (the ARL for a mean shift of size 0) increases from 500 to 2020. The ARL increase is more moderate, but still substantial, for larger mean shifts.

Table 5 ARL values for various size mean shifts for the ARMA(1,1) example when the ARMA parameters coincide with their estimates.

		mean shift magnitude (in units of σ_a)					
chart	control limits	0	1	2	3	4	5
EWMA ($\lambda = 0.1$)	0.202 (standard)	500	101	23.8	8.11	3.54	2.22
EWMA ($\lambda = 0.1$)	0.237 (worst-case)	2020	247	43.3	13.3	5.29	2.89
Shewhart	0.967 (standard)	500	366	168	49.1	7.83	1.38

As another example, with consideration of uncertainty in σ_a, suppose that the parameters of an AR(1) process are estimated using $N = 400$ observations and that the estimates are $\hat{\phi} = 0.5$ and $\hat{\sigma}_a^2 = 1.0$. If $\lambda = 0.1$ and a desired $\text{ARL}_d = 500$ are chosen, again $L = 2.814$. Since $\hat{\sigma}_z = \hat{\sigma}_a (1-v)^{1/2}(1+v)^{-1/2} = 0.2294$, the standard control limits (3.3) are $\pm L \hat{\sigma}_z = \pm 0.646$. Using (A5), the parameter covariance is

$$\Sigma_\gamma \cong \frac{1}{N}\begin{bmatrix} 1-\hat{\phi}^2 & 0 \\ 0 & 2\hat{\sigma}_a^4 \end{bmatrix} = \begin{bmatrix} 1.88 & 0 \\ 0 & 5.0 \end{bmatrix} \times 10^{-3}.$$

If $\alpha = 0.1$ is selected, then (3.6) and (3.7) result in $\hat{V} = [-3.27 \;\; -1.00]^T$, and $\sigma_{z,\alpha} = 0.2516$. The worst-case control limits (3.4) are therefore $\pm L\sigma_{z,\alpha} = \pm 0.708$, which are roughly 10% wider than the standard control limits.

Figure 8 shows results for the AR(1) example that are analogous to Figure 7. Figures 8(a) and 8(c) show the in-control ARL contours as a function of ϕ and σ_a^2 for the standard and worst-case EWMA control limits, respectively. As in the Bayesian analysis of the previous chapter, suppose that the posterior distribution of γ is

approximately multivariate normal with mean $\hat{\gamma}$ and covariance $\hat{\Sigma}_\gamma$. With the worst-case control limits, the probability that the ARL is less than 500 is roughly 0.105, almost identical to the selected value of α.

Figures 8(b) and 8(d) show the corresponding out-of-control ARL contours for a mean shift of magnitude $2\hat{\sigma}_a$. Table 6 gives the out-of-control ARL values for other mean shifts when the true parameters coincide with their estimates. Since the control limits are widened by a lesser extent than in the previous ARMA(1,1) example, the worst-case design results in a much less severe increase in the out-of-control ARL. For mean shifts with magnitude $2\hat{\sigma}_a$ or larger, the out-of-control ARLs increase by roughly 15%, whereas the in-control ARL doubles.

Given the decreased power of the chart that results from widening the control limits, to what extent (or even whether) they should be widened to account for model uncertainty would ideally depend on the costs associated with false alarms and the costs of failing to detect out-of-control conditions, as well as the a priori probability of occurrence of out-of-control conditions. If the costs of false alarms are small, then it may not be desirable to widen the control limits. In the author's experience, however, the costs of frequent false alarms are often quite high when the hidden costs of unnecessary shutdowns and production delays and operators who begin to ignore all alarms, including those that signal real out-of-control conditions, are ocnsidered. To lessen the severity of the tradeoffs in using worst-case control limits, the best solution (when possible) would be to collect a larger sample of data to reduce the parameter
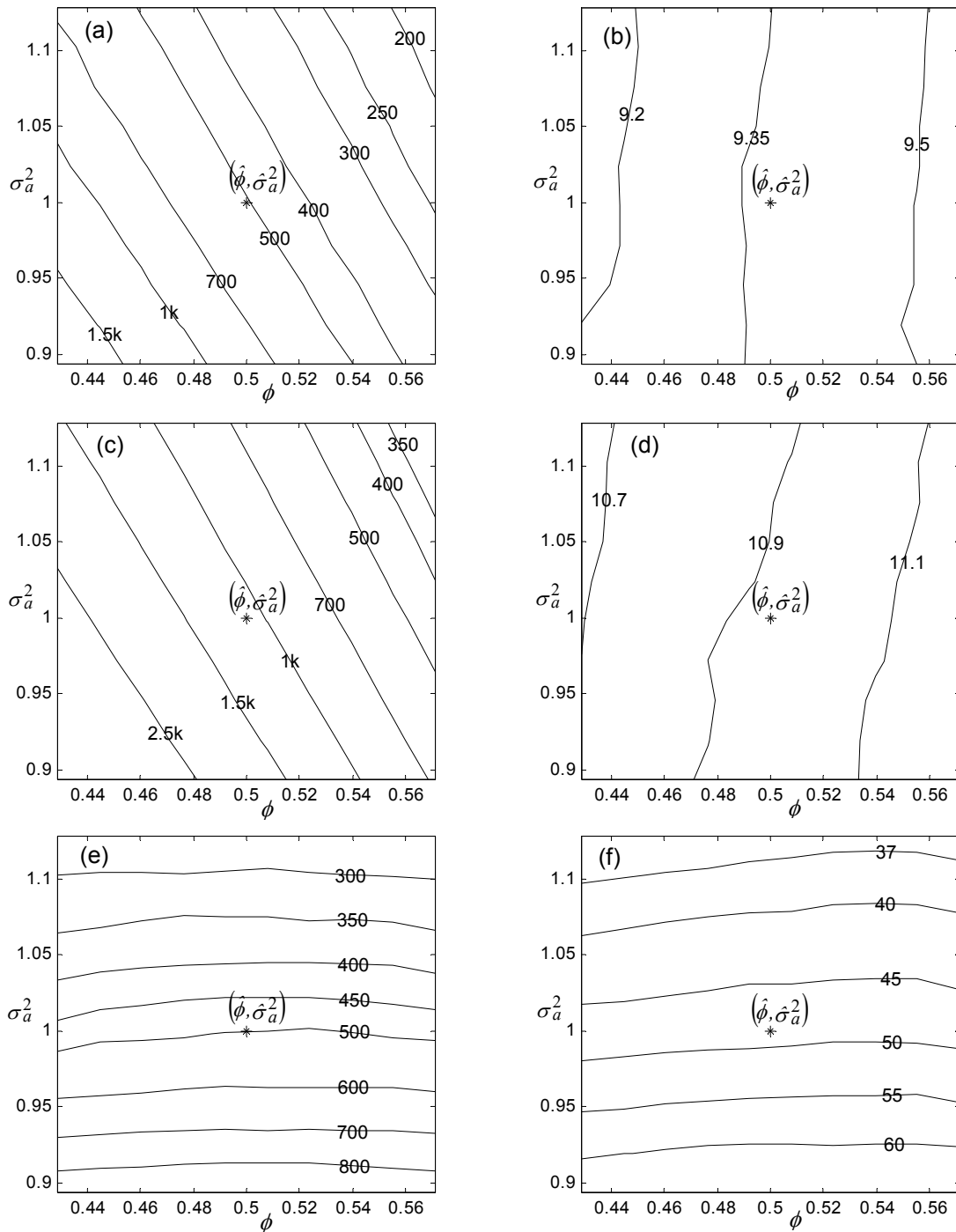
Figure 8 ARL contours as a function of ϕ and σ_a^2 for the AR(1) example. Panels (a), (c), and (e) show the in-control ARLs for the standard EWMA, worst-case EWMA, and Shewhart chart, respectively. Panels (b), (d), and (f) show the out-of-control ARLs for the three charts when the mean shift magnitude is $2\hat{\sigma}_a$.

Table 6 ARL values for various size mean shifts for the AR(1) example when the ARMA parameters coincide with their estimates.

chart	control limits	mean shift magnitude (in units of σ_a)					
		0	1	2	3	4	5
EWMA ($\lambda = 0.1$)	0.646 (standard)	500	30.0	9.37	4.96	3.24	2.34
EWMA ($\lambda = 0.1$)	0.708 (worst-case)	1080	39.6	10.9	5.66	3.68	2.65
Shewhart	3.09 (standard)	500	199	48.1	10.6	2.32	1.10

uncertainty. Guidelines for sample size selection are discussed in Chapter III.4.4.

III.4.3 Shewhart Individual Charts versus EWMA Charts

Figures 7(e) and 7(f) show the in-control and out-of-control ARL contours for a Shewhart individual chart on the residuals in the previous ARMA(1,1) example. Standard control limits of $\pm 3.09 \hat{\sigma}_a = \pm 0.967$ were used, which provide an in-control ARL of 500 when there are no parameter errors. The mean shift magnitude for Figure 7(f) was $3 \hat{\sigma}_a$, the same as for Figures 7(b) and 7(d). The ARL of the Shewhart chart is much less dependent on ϕ and θ than the ARL of an EWMA chart with small λ, because, unlike an EWMA, the Shewhart chart considers only individual residuals and does not take a weighted average of successive residuals. Consequently, residual autocorrelation has little effect on the Shewhart ARL if no supplementary run rules are used. Although an increase in the variance of the residuals will affect the Shewhart ARL, σ_a^2 was assumed equal to $\hat{\sigma}_a^2$ in this example, and small variations in ϕ and θ do not substantially increase the residual variance. Figures 8(e) and 8(f) show

analogous results for a Shewhart chart applied to the residuals in the AR(1) example, where $\pm\ 3.09\ \hat{\sigma}_a$ control limits were again used. In this example, variations in σ_a^2 were also considered. Figure 8(e) shows that the in-control ARL depends predominantly on σ_a^2 and is nearly independent of ϕ over the range of values considered.

Given the relative insensitivity of the Shewhart individual chart with respect to parameter errors, an alternative to using an EWMA with worst-case control limits is to simply use a Shewhart chart with standard control limits. Since the out-of-control ARL for the EWMA is increased when its control limits are widened, one may speculate that the Shewhart chart with standard control limits could provide better detection of mean shifts. Tables 5 and 6 indicate that this is true only for large mean shifts in the examples considered. Even when the worst-case control limits are used, the EWMA still has substantially shorter out-of-control ARLs than the Shewhart chart for small to moderate mean shifts. Table 5 shows that for the ARMA(1,1) example, the Shewhart chart does not surpass the worst-case EWMA in power until the mean shift is between $4\hat{\sigma}_a$ and $5\hat{\sigma}_a$. This is the same level of mean shift at which the Shewhart chart surpasses the EWMA with standard control limits. Table 6 demonstrates similar results for the AR(1) example. Moreover, comparing Figures 7(e) and 8(e) with Figures 7(c) and 8(c), the EWMA with worst-case control limits provides the additional benefit of substantially larger in-control ARLs for most parameter combinations.

III.4.4 Sample Size Requirements

In light of the decreased power that results from widening the EWMA control limits, one may wish to collect a sample of data large enough to ensure $\sigma_{z,\alpha}$ is sufficiently close to $\hat{\sigma}_z$, in which case the worst-case control limits will be close to the standard control limits. It is difficult to provide general guidelines for sample size requirements without some knowledge of the ARMA parameters, since $\sigma_{z,\alpha}$ depends heavily on the parameter estimates. If initial estimates have been obtained from an initial set of data, however, this may be used this to determine how much (or whether) additional data are needed. While waiting for the additional data to be collected, it may be desirable to use both the worst-case and the standard control limits together (refer to Figure 5) as temporary control limits until more accurate parameter estimates and new control limits can be calculated.

Suppose that initial parameter estimates have been obtained and λ and α have been selected. A reasonable strategy is to select the size N of the additional data sample large enough that the resulting percentage difference between $\sigma_{z,\alpha}$ and $\hat{\sigma}_z$ is less than some small value δ (e.g., $\delta = 0.05$). From (3.7), the requirement becomes

$$\frac{\sigma_{z,\alpha}}{\hat{\alpha}_z} = \left[1 + z_\alpha \left(\hat{V}^T \hat{\Sigma}_\gamma \hat{V} \right)^{1/2} \right]^{1/2} < 1 + \delta .$$

Define $\overline{\Sigma}_\gamma = N \hat{\Sigma}_\gamma$. As shown in Appendix A, $\overline{\Sigma}_\gamma$ is a function of the parameter estimates but is otherwise independent of N. If this is substituted into the foregoing inequality, then the sample size requirement reduces to

$$N > \frac{z_\alpha^2 \hat{V}^T \overline{\Sigma}_\gamma \hat{V}}{\delta^2 (2+\delta)^2} . \tag{3.8}$$

To provide some insight into typical sample size requirements, Figure 9 shows contour plots of the required N from Equation (3.8) as a function of $\hat{\phi}$ and $\hat{\theta}$ for an ARMA(1,1) process with four different values of λ. The contour plots are for the specific case of $\delta = 0.05$ and $\alpha = 0.20$. Since neither $\overline{\Sigma}_\gamma$ nor $\hat{V}$ depends on δ and α, results for other δ and α are obtained by multiplying the values of N in Figure 9 by $z_{0.2}^{-2} 0.05^2 2.05^2 z_\alpha^2 \delta^{-2}(2+\delta)^{-2} = 0.0148 z_\alpha^2 \delta^{-2}(2+\delta)^{-2}$. If, for example, a more conservative $\alpha = 0.1$ and the same δ are considered, then the required sample sizes are multiplied by 2.32. For a less conservative $\alpha = 0.3$ and the same δ, the same sizes are multiplied by 0.387. For small δ, (3.8) indicates the required N is approximately inversely proportional to δ^2.

In the ARMA(1,1) example of Chapter III.3 with $\hat{\phi} = 0.87$, $\hat{\theta} = 0.48$, and $N = 197$, the values $\lambda = 0.1$ and $\alpha = 0.1$ were selected. This resulted in worst-case control limits that were 18% wider than the standard control limits. Suppose that one wanted to collect a sample large enough that the worst-case control limits were only 5% wider than the standard limits. From Figure 9, a sample size of approximately 1270 would be required when $\alpha = 0.2$. For $\alpha = 0.1$, the sample size is 2.32 times larger, or $N = 2,940$.
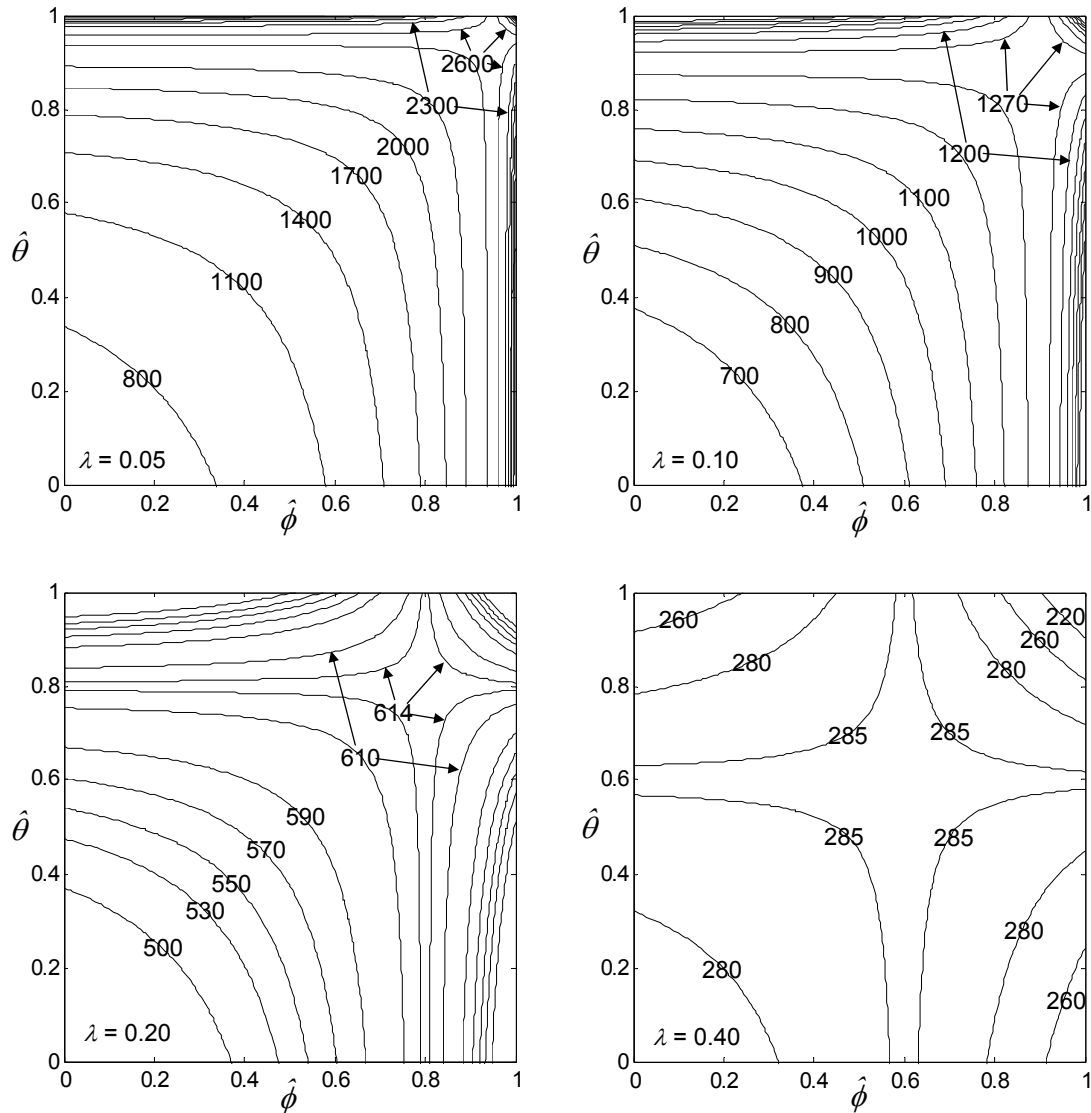
Figure 9 Contours of the required sample size N with $\delta = 0.05$ and $\alpha = 0.20$ for an ARMA(1,1) process with $\lambda = 0.05$, 0.10, 0.20, and 0.40. For other values of δ and α, multiply the contours by $0.0148\, z_\alpha^2\, \delta^{-2}(2+\delta)^{-2}$.

Note that the ridges in Figure 9 are at $\hat{\phi} = 1-\lambda$ and $\hat{\theta} = 1-\lambda$. For a specified λ, the EWMA chart is least robust when the parameter estimates coincide with $1-\lambda$. This does not imply that one should avoid choosing a value of λ that coincides with $1-\hat{\phi}$ or $1-\hat{\theta}$, however. One may show that for any fixed positive values of $\hat{\phi}$ and $\hat{\theta}$, (3.8) increases monotonically as λ decreases.
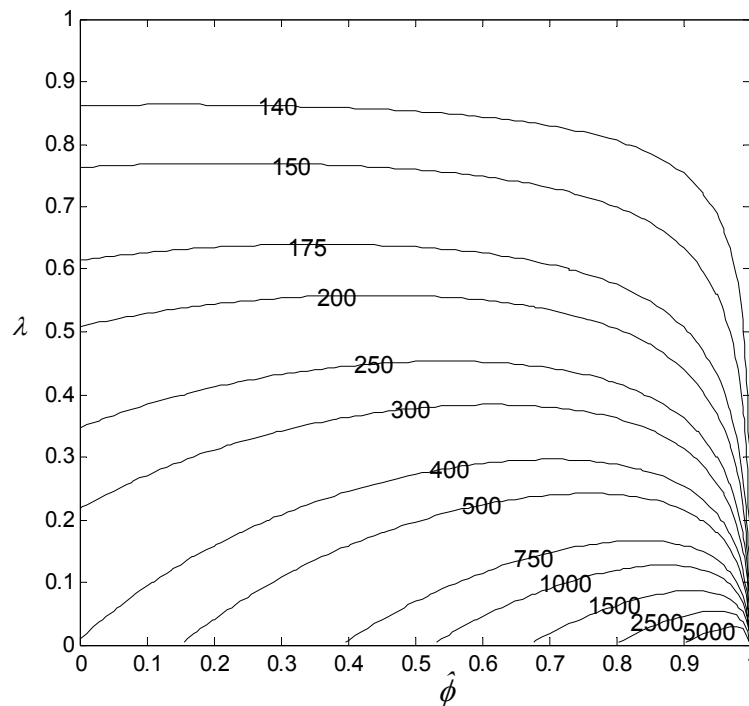
Figure 10 Contours of the required sample size N with $\delta = 0.05$ and $\alpha = 0.20$ for an AR(1) process. For other values of δ and α, multiply the contours by $0.0148\, z_\alpha^2\, \delta^{-2}(2+\delta)^{-2}$. The results for first-order MA and IMA processes are identical if $\hat{\phi}$ is replaced by $\hat{\theta}$.

Figure 10 shows contour plots of the required sample size as a function of $\hat{\phi}$ and λ

for an AR(1) process with $\delta = 0.05$ and $\alpha = 0.20$. Results for other δ and α are again

obtained by multiplying the values of N in Figure 10 by $0.0148\, z_\alpha^2\, \delta^{-2}(2+\delta)^{-2}$. Figure

10 also applies to first-order MA and IMA processes if $\hat{\phi}$ is replaced by $\hat{\theta}$, because of

the symmetry of $\overline{\Sigma}\gamma$ and $\hat{V}$ with respect to the AR and MA parameters. Figures 9 and

10 indicate that very large samples are often required to ensure that $\sigma_{z,\alpha}$ is no more

than 5% larger than $\hat{\sigma}_z$. Even for an AR(1) process with $\alpha = 0.20$, sample sizes close

to 1,000 are required for the typical values $\lambda \approx 0.1$ and $\hat{\phi} > 0.5$.

III.5 Chapter Summary

When designing a residual-based EWMA control chart, a natural measure is to use

wider control limits to account for uncertainty in the estimated parameters. The design

approach of this chapter widens the control limits by an amount commensurate with

the worst-case scenario, in which the ARMA parameters are such that the EWMA

variance equals the maximum value within an appropriate confidence interval.

Assuming an estimate of the parameter covariance matrix is available, or can be

calculated as described in Appendix A, the worst-case design approach involves little

additional complexity relative to the standard design approach.

The disadvantage of widening control limits is the decreased power of control

charts. However, since the purpose of this design method is to prevent an

unacceptably short in-control ARL, the loss of chart power is inevitable. As indicated in simulation results, however, there was no big difference in chart power between the proposed method and the standard EWMA design when a mean shift size is large. In this case, the benefits of using the proposed method are likely to overweigh the loss. Moreover, if the loss by frequent false alarms is considered more serious than the loss by missing a signal of out-of-control, the use of this control limit is recommended although the loss of chart power can be severe in the case of a small mean shift.

CHAPTER IV

ROBUST DESIGN OF RESIDUAL-BASED CONTROL CHARTS FOR AUTOCORRELATED PROCESSES: EXPECTED VARIANCE APPROACH

IV.1 Introduction

In Chapter III, we developed a robust design method for the residual-based EWMA control charts using the worst-case variance. The design method is aimed at guarding against the circumstances when the true ARMA parameters sufficiently differ from their estimates. However, if the estimated and true parameters are such that modeling errors are small enough to be negligible, the resulting control limits are unnecessarily wide and lack sufficient power for detecting mean shifts.

Therefore, this chapter presents another design method for widening the residual-based EWMA control charts to overcome the disadvantages of the worst-case design method. The control limits of this design method are generally widened by a lesser or more suitable amount than in the control limits of the worst-case design approach. As a result, these control limits do not suffer as much from increased out-of-control ARL. To represent model uncertainty, we use a second-order Taylor approximation in this method. Also we use an expected value of the actual EWMA variance instead of the maximum value.

As in the previous chapter, it is assumed that the autocorrelated process data, x_t, follows (1.1) and the in-control process mean has been subtracted so that x_t is 0-mean until there is a shift. The EWMA statistic, z_t, is written as (3.2) when it is applied to the residuals with modeling errors (3.1). When σ_z^2, which is a function of the true parameters for a given λ and ARMA parameter estimates, denotes the actual variance of the EWMA statistic (3.2), the proposed method is to monitor the EWMA statistic (3.2), but to use the control limits

$$\{\text{LCL,UCL}\} = \pm L \sqrt{\text{E}[\sigma_z^2]}$$

instead of the standard control limits (3.3). In Chapter IV.2 we derive the expected variance of an EWMA. Viewing true parameters as random, we use a second-order Taylor approximation to represent the actual EWMA variance and take an expectation on the approximated EWMA variance. The final result is represented by the form of parameter estimates and their covariance matrix. Chapter IV.3 provides results of the expected EWMA variance for special ARMA processes of low orders. From the derived results, we advise actual design procedures in Chapter IV.4. Chapter IV.5 presents some discussions of interest regarding this design method. In the beginning of the chapter, the proposed method is compared to an existing method and the method developed in Chapter III. Afterwards, we discuss differences in the results of expected EWMA variance according to the viewpoint of the random variable—using either true parameters or estimated parameters. Finally, we briefly explain sample size requirements for the proposed method.

IV.2 Expected EWMA Variance

The vector of ARMA parameters is represented as $\gamma = [\phi_1 \; \phi_2 \; ... \; \phi_p \; \theta_1 \; \theta_2 \; ... \; \theta_q]^T$ and the corresponding vector of estimated parameters is represented as $\hat{\gamma}$. The actual EWMA variance, σ_z^2, can be represented by the second-order Taylor approximation about $\gamma = \hat{\gamma}$ as

$$\sigma_z^2 \cong \sigma_z^2 \big|_{\gamma=\hat{\gamma}} + \left[\frac{\partial \sigma_z^2}{\partial \gamma} \big|_{\gamma=\hat{\gamma}} \right]^T (\gamma - \hat{\gamma}) + \frac{1}{2}(\gamma - \hat{\gamma})^T \left[\frac{\partial^2 \sigma_z^2}{\partial \gamma^2} \big|_{\gamma=\hat{\gamma}} \right](\gamma - \hat{\gamma})$$

As in the previous chapter, N denotes the number of observations in the sample. In order to derive an expected EWMA variance, considering parameter uncertainty with the Bayesian view, we use the Bayesian central limit theorem (Carlin and Louis 2000). When the suitable regularity condition holds and the prior of γ is reasonably flat, the posterior of γ can be approximately multivariate normal with mean $\hat{\gamma}$ and covariance $\hat{\Sigma}_\gamma$ for large N. The expected value of the approximated EWMA variance, where the expectation is with respect to the distribution of γ, is written as

$$E[\sigma_z^2] \cong \hat{\sigma}_z^2 + \frac{1}{2} tr \left\{ \frac{\partial^2 \sigma_z^2}{\partial \gamma^2} \big|_{\gamma=\hat{\gamma}} \hat{\Sigma}_\gamma \right\} \tag{4.1}$$

where $\hat{\sigma}_z^2 = \hat{\sigma}_a^2 (1-v)(1+v)^{-1}$ and tr is a matrix trace that sums diagonal elements. The fundamental reason for selecting the Bayesian view is to set a larger impact on the AR parameters in this proposed method. This is will be discussed further in Chapter IV.5.2.

The expected EWMA variance, $E[\sigma_z^2]$, is comprised of two parts. One is the standard EWMA variance and the other is a trace of a matrix, which is a product of the second-order partial derivative matrix of the EWMA variance and the covariance matrix of parameter estimates. The second term can be considered as a gauge that quantifies the level of model uncertainty because the first term is used only for the standard design of EWMA control charts. The second term is represented with a sample size, N, and is inversely proportional to the sample size at the end of the derivation. Thus, the expected EWMA variance can become the standard EWMA variance if we collect plenty of samples to be enough to neglect model uncertainty.

Differentiating (2.2) twice with respect to γ gives

$$\frac{\partial^2 \sigma_z^2}{\partial \gamma^2} = 2\sigma_a^2 \sum_{j=0}^{\infty} \left[g_j \frac{\partial^2 g_j}{\partial \gamma^2} + \left[\frac{\partial g_j}{\partial \gamma}\right]\left[\frac{\partial g_j}{\partial \gamma}\right]^T \right]. \tag{4.2}$$

When (4.2) is substituted for (4.1), the expected EWMA variance is rewritten as

$$E[\sigma_z^2] = \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right) + \hat{\sigma}_a^2 tr\left\{ \sum_{j=0}^{\infty} \hat{g}_j \mathbf{D}_j \hat{\Sigma}_{\gamma} + \sum_{j=0}^{\infty} \mathbf{d}_j \mathbf{d}_j^T \hat{\Sigma}_{\gamma} \right\} \tag{4.3}$$

where we denote

$$\mathbf{D}_j = \frac{\partial^2 g_j}{\partial \gamma^2}\bigg|_{\gamma = \hat{\gamma}}, \text{ and}$$

$$\mathbf{d}_j = \frac{\partial g_j}{\partial \gamma}\bigg|_{\gamma = \hat{\gamma}}.$$

Elements of $\mathbf{D}_j$ are $D_j^{\phi_i,\phi_l}$, $D_j^{\phi_i,\theta_l}$, $D_j^{\theta_i,\phi_l}$, $D_j^{\theta_i,\theta_l}$ and elements of $\mathbf{d}_j$ are $d_j^{\phi_i}$, $d_j^{\theta_l}$.

We derive the form of elements of $\mathbf{D}_j$ in Appendix B and the form of elements of $\mathbf{d}_j$ in Chapter II.2. From Appendix B, the elements of $\mathbf{D}_j$ can be also represented as

$$D_j^{\phi_i,\phi_l} = 2\hat{\Phi}^{-2}(B)\hat{g}_{j-(i+l)} = 2\hat{\Phi}^{-2}(B)\hat{G}(B)\delta_{j-(i+l)} = 2\psi_{j-(i+l)}\left(\hat{\Phi}^{-2}\hat{G}\right) \text{ and}$$

$$D_j^{\phi_i,\theta_l} = -\hat{\Phi}^{-1}(B)\hat{\Theta}^{-1}\hat{g}_{j-(i+l)} = -\hat{\Phi}^{-1}(B)\hat{\Theta}^{-1}(B)\hat{G}(B)\delta_{j-(i+l)}$$

$$= -\psi_{j-(i+l)}\left(\hat{\Phi}^{-1}\hat{\Theta}^{-1}\hat{G}\right).$$

where $\{\psi_j(R): j = 0, 1, 2, \ldots\}$ are the impulse response coefficients of any ARMA transfer function $R(B)$ such that $R(B) = \sum_{j=0}^{\infty} \psi_j(R)B^j$. Similarly, using the impulse response function, $d_j^{\phi_i}$ and $d_j^{\theta_l}$ can be represented as $\psi_{j-i}\left(\hat{\Phi}^{-1}\hat{G}\right)$ and $-\psi_{j-i}\left(\hat{\Theta}^{-1}\hat{G}\right)$ respectively.

Then, from $\sum_{j=0}^{\infty}\hat{g}_j\mathbf{D}_j$ in (4.3), $\sum_{j=0}^{\infty}\hat{g}_j D_j^{\phi_i,\phi_l}$ is

$$\sum_{j=0}^{\infty}\hat{g}_j D_j^{\phi_i,\phi_l} = 2(1-v)\sum_{j=0}^{\infty}v^j\psi_{j-(i+l)}\left(\hat{\Phi}^{-2}\hat{G}\right) = 2(1-v)v^{i+l}\sum_{k=0}^{\infty}\psi_k\left(\hat{\Phi}^{-2}\hat{G}\right)v^k$$

$$= 2(1-v)v^{i+l}\hat{\Phi}^{-2}(v)\hat{G}(v) = 2(1-v)(1+v)^{-1}v^{i+l}\hat{\Phi}^{-2}(v).$$

This follows since $\hat{g}_j = (1-v)v^j$ and $\hat{G}(v) = (1+v)^{-1}$. $\sum_{j=0}^{\infty}\hat{g}_j D_j^{\phi_i,\theta_l}$ becomes

$-(1-v)(1+v)^{-1}v^{i+l}\hat{\Phi}^{-1}(v)\hat{\Theta}^{-1}(v)$ in the same manner.

Therefore, using these in (4.3), $tr\{\sum_{j=0}^{\infty}\hat{g}_j\mathbf{D}_j\hat{\Sigma}_\gamma\}$ becomes

$$tr\{\Sigma_{j=0}^{\infty}\hat{g}_j\mathbf{D}_j\hat{\Sigma}_{\gamma}\}=\frac{1-v}{1+v}tr\left\{\left[\begin{array}{c|c}\dfrac{2\mathbf{V}_p\mathbf{V}_p^T}{\hat{\Phi}^2(v)} & \dfrac{-\mathbf{V}_p\mathbf{V}_q^T}{\hat{\Phi}(v)\hat{\Theta}(v)} \\ \hline \dfrac{-\mathbf{V}_q\mathbf{V}_p^T}{\hat{\Phi}(v)\hat{\Theta}(v)} & \mathbf{0}\end{array}\right]\left[\begin{array}{c|c}\hat{\Sigma}_{\Phi} & \hat{\Sigma}_{\Phi\Theta} \\ \hline \hat{\Sigma}_{\Phi\Theta}^T & \hat{\Sigma}_{\Theta}\end{array}\right]\right\}$$

$$=\frac{2(1-v)}{1+v}\left(\frac{\mathbf{V}_p^T\hat{\Sigma}_{\Phi}\mathbf{V}_p}{\hat{\Phi}^2(v)}-\frac{\mathbf{V}_p^T\hat{\Sigma}_{\Phi\Theta}\mathbf{V}_q}{\hat{\Phi}(v)\hat{\Theta}(v)}\right)$$

where $\hat{\Phi}(v)=1-\hat{\phi}_1 v-\hat{\phi}_2 v^2-\ldots-\hat{\phi}_p v^p$, $\hat{\Theta}(v)=1-\hat{\theta}_1 v-\hat{\theta}_2 v^2-\ldots-\hat{\theta}_q v^q$, $\mathbf{V}_p=[v \quad v^2$

$\ldots v^p]^T$, $\mathbf{V}_q=[v \quad v^2 \ldots v^q]^T$, and $\mathbf{0}$ is a q×q zero matrix. Also, $\hat{\Sigma}_{\Phi}, \hat{\Sigma}_{\Theta}$, and

$\hat{\Sigma}_{\Phi\Theta}$ denote the variance of Φ estimates, the variance of Θ estimates, and the

covariance between Φ estimates and Θ estimates, respectively. These are submatrices

of the matrix $\hat{\Sigma}_{\gamma}$ such that

$$\hat{\Sigma}_{\gamma}=\left[\begin{array}{c|c}\hat{\Sigma}_{\Phi} & \hat{\Sigma}_{\Phi\Theta} \\ \hline \hat{\Sigma}_{\Phi\Theta}^T & \hat{\Sigma}_{\Theta}\end{array}\right].$$

Due to the tedious derivation of $tr\{\Sigma_{j=0}^{\infty}\mathbf{d}_j\mathbf{d}_j^T\hat{\Sigma}_{\gamma}\}$, this is fully shown in Appendix

C. The result is

$$tr\{\Sigma_{j=0}^{\infty}\mathbf{d}_j\mathbf{d}_j^T\hat{\Sigma}_{\gamma}\}$$

$$=\frac{1-v}{(1+v)N}\left[p+q+\frac{2[\hat{\phi}_1 \ 2\hat{\phi}_2 \ 3\hat{\phi}_3 \cdots \mathrm{p}\hat{\phi}_p]\mathbf{V}_p}{\hat{\Phi}(v)}+\frac{2[\hat{\theta}_1 \ 2\hat{\theta}_2 \ 3\hat{\theta}_3 \cdots \mathrm{q}\hat{\theta}_q]\mathbf{V}_q}{\hat{\Theta}(v)}\right].$$

Thus, the expected EWMA variance is

$$E[\sigma_z^2] = \hat{\sigma}_a^2\left(\frac{1-v}{1+v}\right) + \hat{\sigma}_a^2\left(\frac{1-v}{1+v}\right)\left[\frac{2\mathbf{V}_p^T\hat{\mathbf{\Sigma}}_\Phi\mathbf{V}_p}{\hat{\Phi}^2(v)} - \frac{2\mathbf{V}_p^T\hat{\mathbf{\Sigma}}_{\Phi\Theta}\mathbf{V}_q}{\hat{\Phi}(v)\hat{\Theta}(v)} + \frac{p+q}{N}\right.$$

$$+ \frac{2\left[\hat{\phi}_1\; 2\hat{\phi}_2\; 3\hat{\phi}_3\; \cdots\; p\hat{\phi}_p\right]\mathbf{V}_p}{N\hat{\Phi}(v)} + \frac{2\left[\hat{\theta}_1\; 2\hat{\theta}_2\; 3\hat{\theta}_3\; \cdots\; q\hat{\theta}_q\right]\mathbf{V}_q}{N\hat{\Theta}(v)}\left.\right]$$

$$= \hat{\sigma}_a^2\left(\frac{1-v}{1+v}\right)\left[\; 1 + \frac{1}{N}\left[\frac{2\mathbf{V}_p^T\hat{\bar{\mathbf{\Sigma}}}_\Phi\mathbf{V}_p}{\hat{\Phi}^2(v)} - \frac{2\mathbf{V}_p^T\hat{\bar{\mathbf{\Sigma}}}_{\Phi\Theta}\mathbf{V}_q}{\hat{\Phi}(v)\hat{\Theta}(v)} + p+q\right.\right.$$

$$+ \frac{2\left[\hat{\phi}_1\; 2\hat{\phi}_2\; 3\hat{\phi}_3\; \cdots\; p\hat{\phi}_p\right]\mathbf{V}_p}{\hat{\Phi}(v)} + \frac{2\left[\hat{\theta}_1\; 2\hat{\theta}_2\; 3\hat{\theta}_3\; \cdots\; q\hat{\theta}_q\right]\mathbf{V}_q}{\hat{\Theta}(v)}\left.\left.\right]\;\right] \quad (4.6)$$

where $\hat{\bar{\mathbf{\Sigma}}}_\Phi = N\hat{\mathbf{\Sigma}}_\Phi$ and $\hat{\bar{\mathbf{\Sigma}}}_{\Phi\Theta} = N\hat{\mathbf{\Sigma}}_{\Phi\Theta}$.

IV.3 Results for Low-order ARMA Processes

To calculate the expected EWMA variance in (4.6), we need the model order, estimated parameters and their covariance matrix. Box et al. (1994) presented the method for finding the covariance matrix of parameter estimates for general ARMA (*p*,*q*) processes. For special cases of AR (1), AR (2), MA (1), MA (2) and ARMA (1,1), the closed forms of the covariance matrix exist. Subsequently, closed-form expressions of the expected EWMA variance can be determined for these cases. For ARMA (1,1) processes, since the covariance matrix has the form of

$$\hat{\bar{\mathbf{\Sigma}}}_\gamma = \frac{\left(1 - \hat{\phi}_1\hat{\theta}_1\right)}{\left(\hat{\phi}_1 - \hat{\theta}_1\right)^2}\left[\begin{matrix}\frac{\left(1-\hat{\phi}_1^2\right)\left(1-\hat{\phi}_1\hat{\theta}_1\right)}{\left(1-\hat{\phi}_1^2\right)\left(1-\hat{\theta}_1^2\right)} & \frac{\left(1-\hat{\phi}_1^2\right)\left(1-\hat{\theta}_1^2\right)}{\left(1-\hat{\theta}_1^2\right)\left(1-\hat{\phi}_1\hat{\theta}_1\right)}\end{matrix}\right],$$

the expected EWMA variance is,

$\mathrm{E}[\sigma_z^2]$

$$= \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right)[\ 1 + \frac{1}{N}\left[\frac{2v^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)^2\left(1-\hat{\phi}_1^2\right)}{\left(1-\hat{\phi}_1 v\right)^2\left(\hat{\phi}_1-\hat{\theta}_1\right)^2} - \frac{2v^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)\left(1-\hat{\phi}_1^2\right)\left(1-\hat{\theta}_1^2\right)}{\left(1-\hat{\phi}_1 v\right)\left(1-\hat{\theta}_1 v\right)\left(\hat{\phi}_1-\hat{\theta}_1\right)^2}\right.$$

$$\left. +1+1+\frac{2\hat{\phi}_1 v}{1-\hat{\phi}_1 v} + \frac{2\hat{\theta}_1 v}{1-\hat{\theta}_1 v}\right]\ ]$$

$$= \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right)\left[1+\frac{1}{N}\left[\frac{2v^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)\left(1-\hat{\phi}_1^2\right)\left(v-\hat{\theta}_1\right)+2\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1 v\right)\left(1-\hat{\phi}_1\hat{\theta}_1 v^2\right)}{\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1 v\right)^2\left(1-\hat{\theta}_1 v\right)}\right]\right] \quad (4.7)$$

Since the covariance matrix of AR (2) processes is

$$\hat{\bar{\Sigma}}_\gamma = \begin{bmatrix} 1-\hat{\phi}_2^2 & -\hat{\phi}_1\left(1+\hat{\phi}_2\right) \\ -\hat{\phi}_1\left(1+\hat{\phi}_2\right) & 1-\hat{\phi}_2^2 \end{bmatrix}$$

the expected EWMA variance is

$$\mathrm{E}[\sigma_z^2] = \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right)\left[1+\frac{1}{N}\left[\frac{2-2\hat{\phi}_1 v - 2\hat{\phi}_1 v^3 - 6\hat{\phi}_1\hat{\phi}_2 v^3 - \hat{\phi}_2^2 v^2 + \hat{\phi}_2^2 v^4 + v^2 + v^4}{\left(1-\hat{\phi}_1 v - \hat{\phi}_2 v^2\right)^2}\right]\right]$$

For AR (1) processes, the expected EWMA variance becomes

$$\mathrm{E}[\sigma_z^2] = \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right)\left[1+\frac{1}{N}\left[\frac{1-3\hat{\phi}_1^2 v^2 + 2v^2}{\left(1-\hat{\phi}_1 v\right)^2}\right]\right]$$

because the covariance of AR (1) is $1-\hat{\phi}_1^2$.

The expected EWMA variances for MA (2) and MA (1) processes become

$$\mathrm{E}[\sigma_z^2] = \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right)\left[1+\frac{1}{N}\left[\frac{2+2\hat{\theta}_2 v^2}{1-\hat{\theta}_1 v - \hat{\theta}_2 v^2}\right]\right] \quad \text{and}$$

$$E[\sigma_z^2] = \hat{\sigma}_a^2 \left(\frac{1-v}{1+v}\right)\left[1 + \frac{1}{N}\left[\frac{1+\hat{\theta}_1 v}{1-\hat{\theta}_1 v}\right]\right]$$

respectively.

The result of MA (1) processes is exactly the same as the result provided by Apley (2002). This is because the impulse response function, g_j, is the linear function of $\Theta(B)$ as briefly described in Chapter III.2. If parameter estimates are viewed as random instead of true parameters, g_j will be a linear function of $\hat{\Phi}(B)$. Then, the result of AR processes will be exactly the same as the Apley's. More discussions on this can be found in Chapter IV. 5.2.

IV.4 Design with Expected EWMA Variance

The following procedures introduce the actual design. At first, parameters are estimated from a data set. If the estimated model is one of the five special cases in the previous chapter, only the parameter estimates are substituted for the proper expression according to the model order. Most commercial software packages for time series modeling will produce the model order, parameter estimates and the covariance matrix of parameter estimates automatically. Therefore, the expected EWMA variance with any order of ARMA processes can be calculated from (4.6) even if the closed form of the covariance matrix is unavailable.

The recommendation for selecting L is to use the known information such as the tables of Lucas and Saccucci (1990) with a specific value of λ and an intended value

of in-control ARL. As explained in Chapter III.3, one reason for using previous information is that so many factors are involved in choosing L. Another reason for using the same rule is to exclude effects caused by choosing different L values when the proposed method is compared to other robust design procedures in Chapter IV.4. The compared robust residual-based EWMA design methods use the same L values for a specified level of in-control ARL and λ. So, we can fairly investigate differences between the substituted EWMA variances (which are used in each robust design method) for the standard EWMA variance by using the same rule for selecting L.

To illustrate the design procedure with an example, we reuse the Series A data from Box et al. (1994). The model was ARMA(1,1) and parameter estimates were $\hat{\phi}$ = 0.87, $\hat{\theta}$ = 0.48, and $\hat{\sigma}_a^2$ = 0.098. If we use 0.1 as a value of λ and select 500 as a desired level of in-control ARL, L is 2.814 from the table of Lucas and Saccucci (1990). Then, from (4.7), $\sqrt{\mathrm{E}[\sigma_z^2]}$ is 0.0754 and thus control limits are ±0.212. In Chapter III, the standard and worst-case (α=0.1) control limits were ±0.202 and ±0.239 respectively. The control limits of the proposed method are 5% wider than the standard control limits. The widened extent is much reduced in this design method. For the worst-case design, relative increment to the standard control limits was 18%.

IV.5 Discussions

Some points of interest are discussed about the proposed method. Comparison results among robust design methodologies are mainly presented and discussed. The

average run length (ARL) is used for the performance measure of control charts. As in other chapters, the same procedure of Monte Carlo simulation is applied for calculating single in-control or out-of-control ARLs. Refer to Chapter II.6 for details.

IV.5.1 Comparisons

This chapter evaluates the proposed method via several comparisons. Two different design procedures are employed for comparison to the proposed design method. The first one is the design method developed by Apley (2002) (hereafter A method). This method is similar to the proposed method in the sense that the expected value of EWMA variance is used. The differences lie in the viewpoints of random variables and the accuracy level of approximation. Parameter estimates are considered as random and a first-order Taylor approximation is used for representing expected EWMA variance in the A method. The other method for comparison is the design procedure proposed in Chapter III using worst-case EWMA variance (hereafter W method).

Suppose that four sets of parameter estimates are obtained from four separate time series modelings. No modeling errors on σ_a^2 (=1) is assumed for simplicity. For the parameter estimates when λ is 0.05, Table 7 shows control limits (CL) of three robust design approaches and their widened amount to standard control limits. All control limits are designed for providing an in-control ARL of 500. The standard residual-based EWMA control limits are $\pm$ 0.4187. In the W method, α=0.2 (which is the middle value of the recommended range for the significance level) is used to design

the control limits. The relative increment is calculated as $(\sigma - \hat{\sigma}_z)/\hat{\sigma}_z \times 100\%$ where $\hat{\sigma}_z = \hat{\sigma}_a (1-v)^{1/2}(1+v)^{-1/2}$ and σ is the amount that replaces the standard EWMA standard deviation in each robust design approach.

Generally, the control limits of the proposed method are wider than control limits of the A method and narrower than control limits of the W method. Only when $\hat{\theta}_1 = 0.6$ and $N=50$ in Table 7, the control limits of the proposed method are slightly wider than those of the W method. In Table 7, when the sample size is 100, the control limits of the proposed method are around 12 to 17 % wider than those of the standard control limits. For the case of the A method, the relative increment to the standard control limits is below 10%, around 5 to 8 %. The W method provides at least 20 % wider control limits than the standard. Since all these design methods use an EWMA statistic, the reasonable inference is that the proposed method provides larger in-control ARLs than the A method does, whereas the chart power can be less than the power of the A method. On the contrary, the proposed method is expected to provide shorter out-of-control ARLs than those of the W method, as in the case of in-control ARLs. The effects of widening control limits on ARLs are more thoroughly discussed in Chapter III.4.2.

It seems that the control limits of the W method are quite conservative especially when the sample size is relatively large. In Table 7, when the sample size is 500, the relative increment of the W method is approximately three times larger than that of the proposed method. When the sample size is considered, the increased amount to the

standard control limits is substantial. This is due to the inherent characteristic of the W

method, which uses the maximum EWMA variance to guarantee an in-control ARL.

The widened amount of the W method cannot be small unless the sample size and/or

the significance level (α) are quite large. For example, when the sample size is 500,

the control limits of the proposed method are ± 0.4339 at the first set of parameter

estimates. To produce control limits of a similar magnitude in the W method, the

sample size would be around 5000 or α would be larger than 0.3 at least.

Table 7 Control limits of robust EWMA design methods and their increases relative to standard EWMA control limit when λ is 0.05. CL and RI represent control limits and relative increments.

$\hat{\phi_1}$	$\hat{\theta_1}$	N	Proposed		A		W	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	0.5517	31.8%	0.4827	15.2%	0.5484	31.0%
		100	0.4898	17.0%	0.4519	7.9%	0.5138	22.7%
		200	0.4556	8.8%	0.4356	4.0%	0.4879	16.5%
		500	0.4339	3.6%	0.4256	1.6%	0.4637	10.7%
0.9	0.4	50	0.5413	29.3%	0.4775	14.0%	0.5475	30.8%
		100	0.4839	15.6%	0.4491	7.2%	0.5132	22.6%
		200	0.4525	8.1%	0.4342	3.7%	0.4874	16.4%
		500	0.4326	3.3%	0.4250	1.5%	0.4634	10.7%
0.8	0.6	50	0.5455	30.0%	0.4624	10.6%	0.5372	28.3%
		100	0.4863	16.1%	0.4411	5.3%	0.5054	20.7%
		200	0.4538	8.4%	0.4301	2.7%	0.4816	15.0%
		500	0.4331	3.4%	0.4233	1.1%	0.4595	9.7%
0.8	0.4	50	0.5182	23.8%	0.4570	9.1%	0.5339	27.5%
		100	0.4711	12.5%	0.4383	4.7%	0.5029	20.1%
		200	0.4457	6.4%	0.4286	2.4%	0.4798	14.6%
		500	0.4297	2.6%	0.4227	1.0%	0.4583	9.4%

As the sample size increases, the widening in control limits decreases for all cases. The relative increment is influenced by the magnitude of λ as well as the sample size (Apley 2002). As λ increases, the relative increment decreases for every design method. Tables 8 and 9 present information analogous to Table 7 when λ is 0.1 and 0.2 respectively. Overall observations about the results are similar to the case of Table 7. However, relative increments of the proposed method are not as severe as when λ is 0.05. For large N, the conservativeness of the W method is consistent with Table 7. When N is 500, the relative increment of the W method is almost four times to five times larger than that of the proposed method.

Table 8 Control limits of robust EWMA design methods and their increases relative to standard EWMA control limit when λ is 0.1. CL and RI represent control limits and relative increments.

$\hat{\phi}_1$	$\hat{\theta}_1$	N	Proposed		A		W	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	0.7715	19.5%	0.7239	12.1%	0.7958	23.3%
		100	0.7113	10.2%	0.6859	6.2%	0.7549	16.9%
		200	0.6792	5.2%	0.6660	3.2%	0.7246	12.2%
		500	0.6592	2.1%	0.6538	1.3%	0.6966	7.9%
0.9	0.4	50	0.7648	18.5%	0.7169	11.0%	0.7948	23.1%
		100	0.7077	9.6%	0.6821	5.7%	0.7541	16.8%
		200	0.6774	4.9%	0.6641	2.9%	0.7242	12.2%
		500	0.6585	2.0%	0.6531	1.2%	0.6964	7.9%
0.8	0.6	50	0.7753	20.1%	0.7042	9.1%	0.7924	22.8%
		100	0.7134	10.5%	0.6755	4.6%	0.7524	16.5%
		200	0.6803	5.4%	0.6607	2.3%	0.7228	12.0%
		500	0.6597	2.2%	0.6517	0.9%	0.6954	7.7%
0.8	0.4	50	0.7537	16.7%	0.6969	8.0%	0.7910	22.5%
		100	0.7017	8.7%	0.6717	4.0%	0.7513	16.4%
		200	0.6742	4.4%	0.6588	2.0%	0.7219	11.8%
		500	0.6572	1.8%	0.6509	0.8%	0.6948	7.6%

Table 9 Control limits of robust EWMA design methods and their increases relative to standard EWMA control limit when λ is 0.2. CL and RI represent control limits and relative increments.

$\hat{\phi}_1$	$\hat{\theta}_1$	N	Proposed		A		W	
			CL	RI	CL	RI	CL	RI
0.9	0.6	50	1.0889	10.3%	1.0724	8.6%	1.1500	16.5%
		100	1.0394	5.3%	1.0308	4.4%	1.1048	11.9%
		200	1.0137	2.7%	1.0093	2..2%	1.0717	8.5%
		500	0.9980	1.1%	0.9962	0.9%	1.0415	5.5%
0.9	0.4	50	1.0853	9.9%	1.0642	7.8%	1.1485	16.3%
		100	1.0375	5.1%	1.0265	4.0%	1.1037	11.8%
		200	1.0127	2.6%	1.0071	2.0%	1.0709	8.5%
		500	0.9976	1.0%	0.9953	0.8%	1.0410	5.4%
0.8	0.6	50	1.0902	10.4%	1.0579	7.1%	1.1511	16.6%
		100	1.0400	5.3%	1.0232	3.6%	1.1057	12.0%
		200	1.0140	2.7%	1.0054	1.8%	1.0724	8.6%
		500	0.9981	1.1%	0.9946	0.7%	1.0419	5.5%
0.8	0.4	50	1.0820	9.6%	1.0495	6.3%	1.1498	16.5%
		100	1.0358	4.9%	1.0189	3.2%	1.1049	11.9%
		200	1.0118	2.5%	1.0032	1.6%	1.0718	8.6%
		500	0.9972	1.0%	0.9937	0.6%	1.0410	5.4%

To compare the capability of detecting mean shifts among design methods, the chemical process data example in Chapters II and III is used again. We assume that the true parameters equal the estimated parameters and the in-control and out-of-control ARLs are then calculated using Monte Carlo simulation; the results are summarized in Table 10. Two rows, which represent the proposed and A methods, are additionally included in Table 5. If we look at Table 10, the proposed method outperforms the W method, especially when the size of mean shift is small. For the mean shift size of $1\sigma_a$, the proposed method reduces the ARL value to almost half of

that of the W method. For the mean shift size of $2\sigma_a$, the proposed method still outperforms the W method. When the proposed method is compared to the standard design method, differences in timesteps are within one step in detecting the mean shift if the mean shift size is larger than $3\sigma_a$.

As discussed in Chapter III.4.2, widening control limits has an adverse influence on the out-of-control ARLs. Since the control limits of the proposed method are widened by a lesser amount than those of the W method, however, the proposed method does not suffer as much from increased out-of-control ARLs. The A method is in the exact opposite situation when compared to the proposed method. Meanwhile, the proposed method is more likely to provide the desired level of in-control ARL than the A method and W method with the same amount of modeling errors. Consequently, it is reasonable to conclude that the proposed method attains more adequate trade-offs between in-control ARLs and out-of-control ARLs than other design methods.

Table 10 ARL values for various size mean shifts for the ARMA(1,1) example when the ARMA parameters coincide with their estimates.

chart	control limits	mean shift magnitude (in units of σ_a)					
		0	1	2	3	4	5
EWMA ($\lambda = 0.1$)	0.202 (Standard)	500	101	23.8	8.11	3.54	2.22
EWMA ($\lambda = 0.1$)	0.212 (Proposed)	729	129	27.7	9.24	4.00	2.39
EWMA ($\lambda = 0.1$)	0.208 (A method)	612	115	25.5	8.58	3.79	2.30
EWMA ($\lambda = 0.1$)	0.237 (W method)	2020	247	43.3	13.3	5.29	2.89
Shewhart	0.967 (Standard)	500	366	168	49.1	7.83	1.38

IV.5.2 Bayesian or Non-Bayesian

The true parameter vector $\gamma = [\phi_1 \ \phi_2 \ ... \ \phi_p \ \theta_1 \ \theta_2 \ ... \ \theta_q]^T$ is considered as random

in this method. The approximate posterior distribution of γ was used for deriving the

expected EWMA variance in Chapter IV.2. According to the Bayesian central limit

theorem, under the regularity condition and for large N, the posterior distribution of γ

can be approximated as normally distributed with mean of the posterior mode and

covariance of the negative inverse second derivative matrix of the log posterior

evaluated at the mode (Carlin and Louis 2000). In addition, the Bayesian estimation is

approximately equivalent to most estimation methods such as exact or approximate

maximum likelihood and exact or conditional least squares, when the prior

distribution of γ is nearly flat (Box et al. 1994). In this case, it is interpreted that the

posterior mode can be replaced by a general maximum likelihood estimator (MLE).

Thus, the posterior of γ is approximately multivariate normal with mean $\hat{\gamma}$ and

covariance $\hat{\Sigma}_\gamma$.

Within the Bayesian viewpoint, $G(B)$ is the linear function of $\Theta(B)$ as seen in

$G(B) = (1 - v)(1 - vB)^{-1} \hat{\Theta}^{-1}(B) \hat{\Phi}(B) \Phi^{-1}(B) \Theta(B)$. The second-order partial derivatives

of the impulse response function, g_j, with respect to the MA parameters are zero.

Consequently, the expected EWMA variance does not involve the covariance of the

MA parameter estimates. This means we can represent the actual EWMA variance

more accurately by second-order Taylor approximation only in AR and ARMA

processes, and not in MA processes. On the contrary, in non-Bayesian view, $G(B)$ is

the linear function of $\hat{\Phi}(B)$. Then, the second-order differentiation affects MA and ARMA processes in this case. Consequently, the basic reason for using the Bayesian viewpoint is to place greater emphasis on the AR parameters and thus they have more effect on the expected EWMA variance.

IV.5.3 Sample Size Requirements

From (4.6), if we use an infinite number of samples to estimate parameters, the expected EWMA variance goes to the standard EWMA variance. This means that the widened control limits by the amount of model uncertainty can coincide with standard EWMA control limits as long as the sample size is large enough to neglect the effects of model estimation. However, in practice, it is preferred to know a somewhat exact sample size that will guarantee the widened control limits are close to the standard control limits by some degree. If any parameter estimates exist, an additional sample size can be determined to achieve the purpose, which is to know the sample size information (Apley and Lee 2003). Considering that the initial estimation is performed with in-control process data, the additional sample, N, should be collected from the in-control process as well.

When δ denotes the small difference in magnitude between standard EWMA standard deviation, $\hat{\sigma}_z$, and square root of expected EWMA variance, $\sqrt{\mathrm{E}[\sigma_z^2]}$, then the following inequality should be satisfied

$$\frac{\sqrt{\mathrm{E}[\sigma_z^2]}}{\hat{\sigma}_z} \langle 1+\delta$$

For ARMA (1,1) processes, the required sample would be

$$N \rangle \frac{2v^2\left(1-\hat{\phi}_1\hat{\theta}_1\right)\left(1-\hat{\phi}_1^2\right)\left(v-\hat{\theta}_1\right)+2\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1v\right)\left(1-\hat{\phi}_1\hat{\theta}_1v^2\right)}{\left(\delta^2+2\delta\right)\left(\hat{\phi}_1-\hat{\theta}_1\right)\left(1-\hat{\phi}_1v\right)^2\left(1-\hat{\theta}_1v\right)}$$

to obtain the control limits that are $(\delta\times100)\%$ larger than the standard EWMA control limits. From the chemical process data example, the model was ARMA (1,1) and the estimated parameters were $\hat{\phi}_1 = 0.87$, $\hat{\theta}_1 = 0.48$. N should be at least around 310 to ensure that the control limits are 5% larger than the standard control limits when λ is 0.05. When δ is 0.01 with the same λ, the required sample size is around 1600.

For AR (1) processes, sample size requirements should satisfy the following inequality.

$$N \rangle \frac{1-3\hat{\phi}_1^2v^2+2v^2}{\left(\delta^2+2\delta\right)\left(1-\hat{\phi}_1v\right)^2}$$

IV.6 Chapter Summary

Likewise other robust EWMA design procedures, the proposed method modifies the control limits based on the level of model uncertainty. In order to represent the actual EWMA variance, a second-order Taylor approximation is used in the proposed design. This more accurate approximation results in a more fitting increment in modifying control limits by the proposed design. Comparisons to existing design

methods showed suitable properties required for a robust design. In conclusion, the proposed method provides control limits that reduce the risk of excessive false alarms and possess a less severe loss of power in detecting mean shifts.

CHAPTER V

CONCLUSIONS AND FUTURE WORK

V.1 Conclusions

This dissertation has considered model uncertainty in order to develop design procedures that incorporate it into the design process when statistical process control charts are applied to autocorrelated processes. To investigate the effects of modeling errors, Chapter II has represented sensitivity as a function of the autocorrelation of the process. Since the sensitivity is quantified, we have attained an easy interpretation of the effects of modeling errors and compared robustness by numbers between EWMA control charts on x_t and e_t. Especially, the sensitivity of the residual-based EWMA results in simple expressions and used for the robust design of residual-based EWMA control charts in the following chapters.

The main conclusion is that the EWMA on the autocorrelated process data is more sensitive than the EWMA on the residuals with the same λ. Although we would not necessarily use the same λ for both charts, this is important because applying the control charts directly to x_t has been recommended as a more robust alternative to the residual-based control charts with respect to modeling errors. It is also shown that control charts on the feedback controlled output are equally affected by ARMA modeling errors in the same way that residual-based control charts are affected despite the exclusion the modeling errors of β (which is the input/output model parameter in

the closed-loop). Therefore, we can conclude that residual-based control charts are no less robust than widened control charts directly on x_t and control charts on feedback controlled output data in terms of ARMA modeling errors.

Chapter III has developed robust residual-based EWMA control charts using worst-case EWMA variance. The design method widens the control limits by an amount that depends on the level of the model uncertainty. Although some level of robustness is guaranteed with respect to ARMA estimation errors, the inevitable drawback of widening control limits is that the chart power decreases. If failures to detect out-of-control signals are regarded as more critical than false alarms, the worst-case design approach can not be the best option. On the other hand, if the loss by false alarms costs more than the loss by missing a signal, the benefits of using the proposed method are likely to overweigh the loss. To reduce the trade-offs between the in-control ARLs and the out-of-control ARLs, the best answer is to collect large samples when using the proposed method. The guidance for sample size was investigated in the chapter.

Chapter IV has developed another robust residual-based EWMA control charts using expected EWMA variance. This method is intended to overcome the drawback of the worst-case design approach. To represent the actual EWMA variance, this method used more accurate approximation and employed an expected value instead of an maximum value. More precise approximation resulted in a more suitable amount of modification in control limits than in other compared methods. Therefore, this proposed approach could achieve a better balance between false alarms and control

chart power than the existing method and the proposed method in Chapter III. In addition, this method is preferable to worst-case design by virtue of the reduced complexity involved in designing. In this method, the only information that is needed is the parameter estimates and their error covariance matrix. On the contrary, in worst-case design approach we additionally need to choose the significance level. In practice, an additional parameter choice in implementation can be considerable difficulty to users.

V.2 Future Work

Only parameter errors are considered with perfect information of model structure for this dissertation. However, the order of the model (p,q) is often unknown in practical situations. If a Bayesian structure can be employed to address model order uncertainty, a complete robust design method can be developed. For a simple example, suppose that several candidates of model orders exist. If we define the prior distribution of model order as the probability that $\Pr(M=m)=\Pr(x_t$ comes from ARMA (p_m,q_m) process), $\Pr(M=m)$ is the probability for the candidate. Also if the prior of parameters can be determined, necessary information like the posterior distribution of parameters can be derived using a general Bayesian analysis.

Although we have only developed analytical results for EWMA control charts, it is well known (e.g., Adams and Tseng 1998) that other control charts for detecting mean shifts that perform similarly to the EWMA, such as CUSUM charts, are equally sensitive to modeling errors. This brings up the question of whether any control

charting method for autocorrelated processes would be effective at detecting mean shifts, as well as robust to modeling errors. However, the biggest obstacle of developing robust CUSUM design is the structure of the CUSUM statistic, which is analytically intractable. The robust design method for CUSUM charts can be another challenging future work.

Because we have focused on the sensitivity of the EWMA variance, the results are only reflective of the sensitivity of the in-control performance of the control chart. However, a number of empirical studies have shown that the out-of-control performance (e.g., out-of-control ARL) of residual-based charts is also affected by modeling errors (Adams and Tseng 1998; Apley and Shi 1999). Although an analytical analysis of the sensitivity of the out-of-control performance would necessarily involve many factors other than the EWMA variance and would be much more complicated, it would also provide more complete insight into the effects of modeling errors on control charts for autocorrelated data.

REFERENCES

Adams, B. M. and Tseng, I. T. (1998), "Robustness of Forecast-Based Monitoring Schemes," *Journal of Quality Technology*, 30(4), 328-339.

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

Apley, D. W. (2002), "Time Series Control Charts in the Presence of Model Uncertainty," *Journal of Manufacturing Science and Engineering*, 124(4), 891-898.

Apley, D. W. and Shi, J. (1999), "The GLRT for Statistical Process Control of Autocorrelated Processes," *IIE Transactions*, 31(12), 1123-1134.

Apley, D. W. and Lee, H. C. (2003), "Design of Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes with Model Uncertainty," *Technometrics*, 45(3), 187-198.

Åström, K. J. and Wittenmark, B. (1990) *Computer Controlled Systems: Theory and Design, 2nd ed.* Prentice Hall, Englewood Cliffs, NJ.

Berthouex, P. M., Hunter, W. G., and Pallesen, L. (1978), "Monitoring Sewage Treatment Plants: Some Quality Control Aspects," *Journal of Quality Technology*, 10, 139-149.

Box, G., Jenkins, G., and Reinsel, G. (1994), *Time Series Analysis, Forecasting, and Control, 3rd ed.*, Prentice-Hall, Englewood Cliffs, NJ.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods, 2nd ed.*, Springer-Verlag, New York, NY.

Carlin, B. P. and Louis, T. A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, *2nd ed.*, Chapman & Hall/CRC, Boca Raton, FL.

Chen, G. (1997), "The Mean and Standard Deviation of the Run Length Distribution of $\overline{X}$ Charts When Control Limits Are Estimated," *Statistica Sinica*, 7, 789-798.

English, J. R., Krishnamurthi, M., and Sastri, T. (1991), "Quality Monitoring of Continuous Flow Processes," *Computers and Industrial Engineering*, 20(2), 251-260.

Ghosh, B. K., Reynolds, M. R., and Van Hui, Y. (1981), "Shewhart $\overline{X}$-charts with Estimated Process Variance," *Communications in Statistics: Theory and Methods*, 18, 1797-1822.

Johnson, R. A. and Bagshaw, M. (1974), "The Effect of Serial Correlation on the Performance of CUSUM tests," *Technometrics*, 16(1), 103-112.

Jones, L. A. (2002), "The Statistical Design of EWMA Control Charts with Estimated Parameters," *Journal of Quality Technology*, 34(3), 277-288.

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

Lin, W. S. and Adams, B. M. (1996), "Combined Control Charts for Forecast-Based Monitoring Schemes," *Journal of Quality Technology*, 28(3), 289-301.

Lu, C. W. and Reynolds, M. R. (1999), "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes," *Journal of Quality Technology*, 31(2), 166-188.

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

Montgomery, D. C. (2001), *Introduction to Statistical Quality Control, 4rd ed.*, Wiley, New York.

Montgomery, D. C. and Mastrangelo, C. M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data," *Journal of Quality Technology*, 23(3), 179-193.

Montgomery, D. C. and Woodall, W. H. (1997), "A Discussion on Statistically-Based Process Monitoring and Control," *Journal of Quality Technology*, 29(2), 121-162.

Pandit, S. M. and Wu, S. M. (1983), *Time Series and System Analysis with Applications, 1st ed.*, Wiley, New York, NY.

Quesenberry, C. P. (1993), "The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Contol Charts," *Journal of Quality Technology*, 25(4), 237-247.

Runger G. C., Willemain, T. R., and Prabhu, S. (1995), "Average Run Lengths for CUSUM Control Charts Applied to Residuals," *Communications in Statistics: Theory and Methods*, 24(1), 273-282.

Superville, C. R. and Adams, B. M. (1994), "An Evaluation of Forecast-Based Quality Control Schemes," *Communications in Statistics: Simulation and Computation*, 23(3), 645-661.

Vander Wiel, S. A. (1996), "Monitoring Processes That Wander Using Integrated Moving Average Models," *Technometrics*, 38(2), 139-151.

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

Wardell, D. G., Moskowitz, H. and Plante, R. D. (1994), "Run-Length Distributions of Special-Cause Control Charts for Correlated Processes," *Technometrics*, 36(1), 3-17.

Woodall, W. H. and Montgomery, D. C. (1999), "Research Issues and Ideas in Statistical Process Control," *Journal of Quality Technology*, 31(4), 376-386.

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

APPENDIX A

CALCULATING PARAMETER COVARIANCE Σ_γ

Assume the ARMA parameters are estimated using a method based on minimizing the sum of the squares of the model residuals, such as the nonlinear least squares or approximate maximum likelihood methods described in Box, et al. (1994). For sample size N sufficiently large, the parameter covariance matrix is (Box, et al., 1994)

$$\Sigma_\gamma \cong \begin{bmatrix} \Sigma_\eta & \mathbf{0} \\ \mathbf{0}^T & 2N^{-1}\sigma_a^4 \end{bmatrix} = \frac{1}{N} \begin{bmatrix} \sigma_a^2 \Sigma_w^{-1} & \mathbf{0} \\ \mathbf{0}^T & 2\sigma_a^4 \end{bmatrix}, \tag{A1}$$

where $\mathbf{0}$ denotes a column vector of $p+q$ 0s, and Σ_η denotes the covariance of $\boldsymbol{\eta} = [\phi_1$ $\phi_2 \ldots \phi_p \; \theta_1 \; \theta_2 \ldots \theta_q]^T$. The matrix Σ_w is defined as the covariance matrix of the random vector $\mathbf{w}_t = [u_t \, u_{t-1} \ldots u_{t-p} \, v_t \, v_{t-1} \ldots v_{t-q}]^T$, where u_t and v_t are defined via $u_t = \Phi^{-1}(B)a_t$ and $v_t = -\Theta^{-1}(B)a_t$.

To calculate Σ_w, rewrite $u_t = \sum_{j=0}^{\infty} g_{\phi,j} a_{t-j}$ and $v_t = -\sum_{j=0}^{\infty} g_{\theta,j} a_{t-j}$, where the $g_{\phi,j}$'s and $g_{\theta,j}$'s are the impulse response coefficients of $\Phi^{-1}(B)$ and $\Theta^{-1}(B)$, respectively. Note that the impulse response coefficients can be calculated recursively for $j = 1, 2, \ldots$, via

$$g_{\phi,j} = \phi_1 g_{\phi,j-1} + \phi_2 g_{\phi,j-2} + \ldots + \phi_p g_{\phi,j-p}, \text{ and} \tag{A2}$$

$$g_{\theta,j} = \theta_1 g_{\theta,j-1} + \theta_2 g_{\theta,j-2} + \ldots + \theta_q g_{\theta,j-q} \tag{A3}$$

with $g_{\phi,j} = g_{\theta,j} = 0$ for $j < 0$, and $g_{\phi,0} = g_{\theta,0} = 1$. If the matrix

$$\mathbf{H} = \begin{bmatrix} g_{\phi,0} & 0 & \cdots & 0 & -g_{\theta,0} & 0 & \cdots & 0 \\ g_{\phi,1} & g_{\phi,0} & \ddots & \vdots & -g_{\theta,1} & -g_{\theta,0} & \ddots & \vdots \\ g_{\phi,2} & g_{\phi,1} & \ddots & 0 & -g_{\theta,2} & -g_{\theta,1} & \ddots & 0 \\ \vdots & \vdots & & g_{\phi,0} & \vdots & \vdots & & -g_{\theta,0} \\ g_{\phi,p} & g_{\phi,p-1} & \cdots & g_{\phi,1} & -g_{\theta,q} & -g_{\theta,q-1} & \cdots & -g_{\theta,1} \\ g_{\phi,p+1} & g_{\phi,p} & \cdots & g_{\phi,2} & -g_{\theta,q+1} & -g_{\theta,q} & \cdots & -g_{\theta,2} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \end{bmatrix}$$

is constructed from the impulse response coefficients, then $\Sigma_w = \sigma_a^2 \mathbf{H}^T\mathbf{H}$ results, and

$\sigma_a^2 \Sigma_w^{-1} = [\mathbf{H}^T\mathbf{H}]^{-1}$ can be substituted in (A1). Since the impulse response coefficients

decay exponentially for stable, invertible ARMA processes, the number of rows that

are needed in $\mathbf{H}$ will generally be reasonable.

Because the true ARMA parameters are unknown, their estimates must be

substituted into (A1) through (A3) to calculate the estimate $\hat{\Sigma}_\gamma$ for use in the

confidence interval (3.7). Box, et al. (1994) shows that for first-order AR, MA, and

ARMA processes, the estimated covariance of η reduces to the following:

ARMA(1,1): $\qquad \hat{\Sigma}_\eta = \dfrac{\left(1-\hat{\phi}\hat{\theta}\right)}{N\left(\hat{\phi}-\hat{\theta}\right)^2} \begin{bmatrix} \left(1-\hat{\phi}^2\right)\left(1-\hat{\phi}\hat{\theta}\right) & \left(1-\hat{\phi}^2\right)\left(1-\hat{\theta}^2\right) \\ \left(1-\hat{\phi}^2\right)\left(1-\hat{\theta}^2\right) & \left(1-\hat{\theta}^2\right)\left(1-\hat{\phi}\hat{\theta}\right) \end{bmatrix}$ $\qquad$ (A4)

AR(1): $\qquad \hat{\Sigma}_\eta = \dfrac{1-\hat{\phi}^2}{N}$ $\qquad$ (A5)

and

MA(1): $\qquad \hat{\Sigma}_\eta = \dfrac{1-\hat{\theta}^2}{N}$

APPENDIX B

DERIVATIONS OF $D_j^{\phi_i,\phi_l}$, $D_j^{\phi_i,\theta_l}$, $D_j^{\theta_i,\phi_l}$ AND $D_j^{\theta_i,\theta_l}$

From the relationship $g_j = \Phi^{-1}(B)\Theta(B)h_j$, differentiating both sides with respect

to ϕ_i and θ_l gives

$$\frac{\partial g_j}{\partial \phi_i} = g_{j-i} + \sum_{k=1}^{p} \phi_k \frac{\partial g_{j-k}}{\partial \phi_i} \text{ , and} \tag{B1}$$

$$\frac{\partial g_j}{\partial \theta_l} = -h_{j-l} + \sum_{k=1}^{p} \phi_k \frac{\partial g_{j-k}}{\partial \theta_l} \tag{B2}$$

Differentiating both sides of (B1) with respect to ϕ_l gives

$$\frac{\partial^2 g_j}{\partial \phi_i \partial \phi_l} = \frac{\partial}{\partial \phi_l}\left(g_{j-i} + \sum_{k=1}^{p} \phi_k \frac{\partial g_{j-k}}{\partial \phi_i}\right) = \frac{\partial g_{j-i}}{\partial \phi_l} + \sum_{k=1}^{p}\left(\frac{\partial}{\partial \phi_l}\left(\phi_k \frac{\partial g_{j-k}}{\partial \phi_i}\right)\right)$$

$$= \frac{\partial g_{j-i}}{\partial \phi_l} + \sum_{k=1}^{p} \frac{\partial \phi_k}{\partial \phi_l} \frac{\partial g_{j-k}}{\partial \phi_i} + \sum_{k=1}^{p} \phi_k \frac{\partial^2 g_{j-k}}{\partial \phi_i \partial \phi_l}$$

$$= \frac{\partial g_{j-i}}{\partial \phi_l} + \frac{\partial g_{j-l}}{\partial \phi_i} + \sum_{k=1}^{p} \phi_k \frac{\partial^2 g_{j-k}}{\partial \phi_i \partial \phi_l}$$

Therefore, by the definition of $D_j^{\phi_i,\phi_l}$,

$$D_j^{\phi_i,\phi_l} = \frac{\partial^2 g_j}{\partial \phi_i \partial \phi_l}\bigg|_{\gamma=\hat{\gamma}} = \frac{\partial g_{j-i}}{\partial \phi_l}\bigg|_{\gamma=\hat{\gamma}} + \frac{\partial g_{j-l}}{\partial \phi_i}\bigg|_{\gamma=\hat{\gamma}} + \sum_{k=1}^{p} \phi_k \frac{\partial^2 g_{j-k}}{\partial \phi_i \partial \phi_l}\bigg|_{\gamma=\hat{\gamma}}$$

$$\left(1 - \sum_{k=1}^{p} \hat{\phi}_k B^k\right) D_j^{\phi_i,\phi_l} = d_{j-i}^{\phi_l} + d_{j-l}^{\phi_i}$$

$$D_j^{\phi_i,\phi_l} = \frac{d_{j-i}^{\phi_l} + d_{j-l}^{\phi_i}}{\hat{\Phi}(B)} = \frac{2}{\hat{\Phi}^2(B)} \hat{g}_{j-(i+l)}.$$

Differentiating both sides of (B1) with respect to θ_l gives

$$\frac{\partial^2 g_j}{\partial \phi_i \partial \theta_l} = \frac{\partial}{\partial \theta_l}\left(g_{j-i} + \sum_{k=1}^{p} \phi_k \frac{\partial g_{j-k}}{\partial \phi_i}\right) = \frac{\partial g_{j-i}}{\partial \theta_l} + \sum_{k=1}^{p} \phi_k \frac{\partial^2 g_{j-k}}{\partial \phi_i \partial \theta_l}$$

Therefore, by the definition of $D_j^{\phi_i,\theta_l}$

$$D_j^{\phi_i,\theta_l} = \frac{\partial^2 g_j}{\partial \phi_i \partial \theta_l}\bigg|_{\gamma=\hat{\gamma}} = \frac{\partial g_{j-i}}{\partial \theta_l}\bigg|_{\gamma=\hat{\gamma}} + \sum_{k=1}^{p} \phi_k \frac{\partial^2 g_{j-k}}{\partial \phi_i \partial \theta_l}\bigg|_{\gamma=\hat{\gamma}}$$

$$\left(1 - \sum_{k=1}^{p} \hat{\phi}_k B^k\right) D_j^{\phi_i,\theta_l} = d_{j-i}^{\theta_l}$$

$$D_j^{\phi_i,\theta_l} = \frac{d_{j-i}^{\theta_l}}{\hat{\Phi}(B)} = -\frac{1}{\hat{\Phi}(B)\hat{\Theta}(B)} \hat{g}_{j-(i+l)}.$$

It can be shown that $D_j^{\theta_i,\phi_l}$ is exactly same as $D_j^{\phi_i,\theta_l}$.

Differentiating both sides of (B2) with respect to θ_l gives

$$\frac{\partial^2 g_j}{\partial \theta_i \partial \theta_l} = \frac{\partial}{\partial \theta_i}\left(-h_{j-l} + \sum_{k=1}^{p} \phi_k \frac{\partial g_{j-k}}{\partial \theta_l}\right) = \sum_{k=1}^{p} \phi_k \frac{\partial^2 g_{j-k}}{\partial \theta_i \partial \theta_l}$$

Therefore,

$$D_j^{\theta_i,\theta_l} = \frac{\partial^2 g_j}{\partial \theta_i \partial \theta_l}\bigg|_{\gamma=\hat{\gamma}} = 0.$$

APPENDIX C

DERIVATION OF $tr\{\sum_{j=0}^{\infty} \mathbf{d}_j \mathbf{d}_j^T \Sigma_{\gamma}\}$

An approximate expression (that is asymptotically exact) for the covariance matrix of $\hat{\gamma}$ is (Box et al. 1994)

$$\Sigma_{\gamma} = \frac{\sigma_a^2}{N} \Sigma_w^{-1} \tag{C1}$$

where Σ_w is the covariance matrix of the random vector $\mathbf{w}_t$, defined as $\mathbf{w}_t = [\ u_t\ u_{t-1}$

$\dots\ u_{t-p+1}\ v_t\ v_{t-1}\ \dots v_{t-q+1}\ ]^T$. The random processes u_t and v_t are defined as $u_t = \hat{\Phi}^{-1}(B)a_t$ and $v_t = -\hat{\Theta}^{-1}(B)a_t$.

Let $\mathbf{y}_t = (1-v)(1-vB)^{-1}\mathbf{w}_t = (1-v)\sum_{k=0}^{\infty} v^k \mathbf{w}_{t-k}$, and note that the elements of $\mathbf{y}_t$ are time-delayed versions of $(1-v)(1-vB)^{-1}\hat{\Phi}^{-1}(B)a_t$ and $-(1-v)(1-vB)^{-1}\hat{\Theta}^{-1}(B)a_t$.

Therefore, since $d_j^{\phi_i}$ and $d_j^{\theta_l}$ are the delayed impulse responses of the filters $(1-v)(1-vB)^{-1}\hat{\Phi}^{-1}(B)$ and $-(1-v)(1-vB)^{-1}\hat{\Theta}^{-1}(B)$, it follows that

$$\Sigma_y = \sigma_a^2 \sum_{j=0}^{\infty} \mathbf{d}_j \mathbf{d}_j^T \ \text{ or } \ \sum_{j=0}^{\infty} \mathbf{d}_j \mathbf{d}_j^T = \frac{1}{\sigma_a^2} \Sigma_y. \tag{C2}$$

We can also write

$$\Sigma_y = \mathrm{E}[\ \mathbf{y}_t \mathbf{y}_t^T\ ] = (1-v)^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} v^j v^k \ \mathrm{E}[\ \mathbf{w}_{t-j} \mathbf{w}_{t-k}^T\ ]$$

Therefore, from (C1) and (C2)

$$tr\{\sum_{j=0}^{\infty}\mathbf{d}_j\mathbf{d}_j^T\Sigma_{\boldsymbol{\gamma}}\} = tr\{\frac{1}{\sigma_a^2}\Sigma_{\mathbf{y}}\frac{\sigma_a^2}{N}\Sigma_{\mathbf{w}}^{-1}\}$$

$$= \frac{1}{N}\ tr\{\Sigma_{\mathbf{y}}\Sigma_{\mathbf{w}}^{-1}\}$$

$$= \frac{(1-v)^2}{N}\ tr\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}v^jv^k\ \mathrm{E}[\ \mathbf{w}_{t-j}\mathbf{w}_{t-k}^T\ ]\Sigma_{\mathbf{w}}^{-1}\}$$

$$= \frac{(1-v)^2}{N}\ tr\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}v^jv^k\ \mathrm{E}[\ \mathbf{w}_t\mathbf{w}_{t-|k-j|}^T\ ]\Sigma_{\mathbf{w}}^{-1}\} \qquad \text{(C3)}$$

To evaluate this, $\mathbf{w}_t$ is written as

$$\mathbf{w}_t = \mathbf{A}\mathbf{w}_{t-1} + \mathbf{b}a_t$$

where $\mathbf{A} = \begin{bmatrix} \hat{\phi}_1 & \hat{\phi}_2 & \cdots & \cdots & \hat{\phi}_p & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & & & 0 & 0 & 0 & & & \vdots \\ 0 & 1 & 0 & & \vdots & \vdots & & \ddots & & 0 \\ \vdots & & \ddots & \ddots & 0 & \vdots & & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \hdashline 0 & 0 & \cdots & \cdots & 0 & \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \cdots & \hat{\theta}_q \\ 0 & 0 & & & \vdots & 1 & 0 & & & 0 \\ \vdots & & \ddots & & 0 & 0 & 1 & 0 & & \\ \vdots & & & \ddots & \vdots & \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ \hdashline -1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$

Then for any j, k, we have

$$\mathbf{w}_t = \mathbf{A}^{|k-j|}\mathbf{w}_{t-|k-j|} + \text{function of } \{a_t\ a_{t-1}\ \dots\ a_{t-|k-j|+1}\}$$

and the function of $\{a_t\ a_{t-1}\ \dots\ a_{t-|k-j|+1}\}$ is independent of $\mathbf{w}_{t-|k-j|}$.

Therefore,

$$E[\mathbf{w}_t \mathbf{w}^T_{t-|k-j|}] = \mathbf{A}^{|k-j|}\boldsymbol{\Sigma}_{\mathbf{w}}.$$ (C4)

When (C4) is combined with (C3),

$$tr\{\Sigma^\infty_{j=0}\mathbf{d}_j\mathbf{d}^T_j\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\} = \frac{(1-v)^2}{N}\ tr\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}v^jv^k\ \mathbf{A}^{|k-j|}\}$$

$$= \frac{(1-v)^2}{N}\ tr\{2\sum_{j=0}^{\infty}\sum_{k=j}^{\infty}v^jv^k\ \mathbf{A}^{|k-j|} - \sum_{j=0}^{\infty}v^{2j}\ \mathbf{I}\}$$

$$= \frac{(1-v)^2}{N}\ [2tr\{\sum_{j=0}^{\infty}v^{2j}\sum_{l=0}^{\infty}(v\mathbf{A})^l\} - (p+q)\sum_{j=0}^{\infty}v^{2j}]$$

$$= \frac{(1-v)^2}{(1-v^2)N}\ [2tr\{\sum_{l=0}^{\infty}(v\mathbf{A})^l\} - (p+q)]$$

$$= \frac{(1-v)}{(1+v)N}\ [2tr\{[\mathbf{I}-v\mathbf{A}]^{-1}\} - (p+q)]$$ (C5)

where $\mathbf{I}$ is a $(p+q)\times(p+q)$ identity matrix.

The matrix, $[\mathbf{I}-v\mathbf{A}]^{-1}$, in (C5) is investigated only by diagonal elements because of the trace operator. The method for finding diagonal elements of the matrix $[\mathbf{I}-v\mathbf{A}]^{-1}$ is provided in Appendix D.

As shown in Appendix D, the 1st and ith diagonal elements of the matrix, $[\mathbf{I}-v\mathbf{A}_\Phi]^{-1}$ are $1/\hat{\Phi}(v)$ and $\left(1-\sum_{j=1}^{i-1}\hat{\phi}_jv^j\right)/\hat{\Phi}(v)$ respectively where $2 \leq i \leq p$. The 1st and ith diagonal elements of the matrix, $[\mathbf{I}-v\mathbf{A}_\Theta]^{-1}$ are $1/\hat{\Theta}(v)$ and $\left(1-\sum_{j=1}^{i-1}\hat{\theta}_jv^j\right)/\hat{\Theta}(v)$ respectively where $2 \leq i \leq q$.

Therefore,

$$2tr\{ [\mathbf{I}-\nu\mathbf{A}]^{-1}\} - (p+q)$$

$$= \frac{2\left\{1 + \left(1-\hat{\phi}_1\nu\right) + \left(1-\hat{\phi}_1\nu-\hat{\phi}_2\nu^2\right) + \cdots + \left(1-\hat{\phi}_1\nu-\hat{\phi}_2\nu^2-\cdots-\hat{\phi}_{p-1}\nu^{p-1}\right)\right\}}{\hat{\Phi}(\nu)} - p$$

$$+ \frac{2\left\{1 + \left(1-\hat{\theta}_1\nu\right) + \left(1-\hat{\theta}_1\nu-\hat{\theta}_2\nu^2\right) + \cdots + \left(1-\hat{\theta}_1\nu-\hat{\theta}_2\nu^2-\cdots-\hat{\theta}_{q-1}\nu^{q-1}\right)\right\}}{\hat{\Theta}(\nu)} - q$$

$$= \frac{p\left\{1-\hat{\phi}_1\nu-\cdots-\hat{\phi}_{p-1}\nu^{p-1}-\hat{\phi}_p\nu^p\right\} + 2\left\{\hat{\phi}_1\nu + 2\hat{\phi}_2\nu^2 + \cdots + (p-1)\hat{\phi}_{p-1}\nu^{p-1} + p\hat{\phi}_p\nu^p\right\}}{\hat{\Phi}(\nu)}$$

$$+ \frac{q\left\{1-\hat{\theta}_1\nu-\cdots-\hat{\theta}_{q-1}\nu^{q-1}-\hat{\theta}_q\nu^q\right\} + 2\left\{\hat{\theta}_1\nu + 2\hat{\theta}_2\nu^2 + \cdots + (q-1)\hat{\theta}_{q-1}\nu^{q-1} + q\hat{\theta}_q\nu^q\right\}}{\hat{\Theta}(\nu)}$$

$$= p + q + \frac{2\left[\hat{\phi}_1\ 2\hat{\phi}_2\ 3\hat{\phi}_3\ \cdots\ p\hat{\phi}_p\right]\mathbf{V}_p}{\hat{\Phi}(\nu)} + \frac{2\left[\hat{\theta}_1\ 2\hat{\theta}_2\ 3\hat{\theta}_3\ \cdots\ q\hat{\theta}_q\right]\mathbf{V}_q}{\hat{\Theta}(\nu)}$$

Finally,

$$tr\{\textstyle\sum_{j=0}^{\infty}\mathbf{d}_j\mathbf{d}_j^T\mathbf{\Sigma}_\gamma\} =$$

$$\frac{(1-\nu)}{(1+\nu)N}\left[p + q + \frac{2\left[\hat{\phi}_1\ 2\hat{\phi}_2\ 3\hat{\phi}_3\ \cdots\ p\hat{\phi}_p\right]\mathbf{V}_p}{\hat{\Phi}(\nu)} + \frac{2\left[\hat{\theta}_1\ 2\hat{\theta}_2\ 3\hat{\theta}_3\ \cdots\ q\hat{\theta}_q\right]\mathbf{V}_q}{\hat{\Theta}(\nu)} \right]$$

APPENDIX D

FINDING DIAGONAL ELEMENTS OF MATRIX $[\mathbf{I}-v\mathbf{A}]^{-1}$

The matrix $\mathbf{A}$ in Appendix C can be partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_\Phi & 0 \\ 0 & \mathbf{A}_\Theta \end{bmatrix} \text{ and}$$

let e_i denote a column vector that has 1 for the ith row element and zeros for all other

elements such that $e_i = [\, 0\, 0\, \dots 0\, 1\, 0\, \dots 0]^T$.

The submatrix of $\mathbf{A}$, $\mathbf{A}_\Phi$, satisfies the following two properties.

$$e_i^T \mathbf{A}_\Phi = \sum_{j=1}^{p} \hat{\phi}_j e_j^T , \quad \text{for} \quad i = 1 \tag{D1}$$

$$e_i^T \mathbf{A}_\Phi = e_{i-1}^T , \qquad \text{for } 2 \le i \le p \tag{D2}$$

and the column vector e_i satisfies

$$e_i^T \mathbf{I} e_j = 1 \qquad \text{for} \quad i = j \tag{D3}$$

$$e_i^T \mathbf{I} e_j = 0 \qquad \text{for} \quad i \ne j \tag{D4}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix.

Denote $[\mathbf{I}-v\mathbf{A}_\Phi]^{-1}$ as $\mathbf{M}$ for notational convenience. The matrix $\mathbf{M}$ is written as

$$\mathbf{M} = \left[\mathbf{I} - v\mathbf{A}_\Phi\right]^{-1} = \sum_{k=0}^{\infty} \left(v\mathbf{A}_\Phi\right)^k = \mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots.$$

Using (D1)–(D4), for $2 \le i \le p$, the ith diagonal element of $\mathbf{M}$ is

$$\mathbf{M}_{ii} = e_i^T \mathbf{M}\, e_i$$

$$= e_i^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$= e_i^T \mathbf{I}\, e_i + v e_i^T \mathbf{A}_\Phi \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$= 1 + v e_{i-1}^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$= 1 + v \left\{ e_{i-1}^T \mathbf{I}\, e_i + v e_{i-1}^T \mathbf{A}_\Phi \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i \right\}$$

$$= 1 + v^2 e_{i-2}^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$= 1 + v^2 \left\{ e_{i-2}^T \mathbf{I}\, e_i + v e_{i-2}^T \mathbf{A}_\Phi \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i \right\}$$

$$= 1 + v^3 e_{i-3}^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$\vdots$$

$$= 1 + v^{i-1} \left\{ e_1^T \mathbf{I}\, e_i + v e_1^T \mathbf{A}_\Phi \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i \right\}$$

$$= 1 + v^i \left(\hat{\phi}_1 e_1^T + \hat{\phi}_2 e_2^T + \cdots\cdots + \hat{\phi}_p e_p^T \right) \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$= 1 + v^i \hat{\phi}_1 e_1^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$+ v^i \hat{\phi}_2 e_2^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$\vdots$$

$$+ v^i \hat{\phi}_i e_i^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

$$\vdots$$

$$+ v^i \hat{\phi}_p e_p^T \left(\mathbf{I} + v\mathbf{A}_\Phi + v^2 \mathbf{A}_\Phi^2 + v^3 \mathbf{A}_\Phi^3 + \cdots\cdots \right) e_i$$

At the last equation above,

$$e_j^T \mathbf{M}\, e_i = v^{j-i}\left(e_i^T \mathbf{M}\, e_i - 1\right) \quad : j < i$$

$$v^{j-i}\left(e_i^T \mathbf{M}\, e_i\right) \quad : i \leq j \leq p. \tag{D5}$$

Using these, $\mathbf{M}_{ii}$ becomes

$$\mathbf{M}_{ii} = 1 + v^i \sum_{j=1}^{i-1} \hat{\phi}_j v^{j-i}\left(e_i^T \mathbf{M}\, e_i - 1\right) + v^i \sum_{j=i}^{p} \hat{\phi}_j v^{j-i} e_i^T \mathbf{M}\, e_i$$

$$= 1 + v^i \sum_{j=1}^{p} \hat{\phi}_j v^{j-i} e_i^T \mathbf{M}\, e_i - v^i \sum_{j=i}^{i-1} \hat{\phi}_j v^{j-i}$$

$$= 1 + \sum_{j=1}^{p} \hat{\phi}_j v^{j} e_i^T \mathbf{M}\, e_i - \sum_{j=i}^{i-1} \hat{\phi}_j v^{j}$$

Thus, the ith element of the matrix, $[\mathbf{I} - v\mathbf{A}_\phi]^{-1}$, is

$$\mathbf{M}_{ii} = \frac{1 - \sum_{j=1}^{i-1} \hat{\phi}_j v^{j}}{1 - \sum_{j=1}^{p} \hat{\phi}_j v^{j}} = \frac{1 - \sum_{j=1}^{i-1} \hat{\phi}_j v^{j}}{\hat{\Phi}(v)}$$

The first diagonal element of $\mathbf{M}$, $\mathbf{M}_{11}$, can be also obtained similarly. In this case instead of (D5), $e_j^T \mathbf{M}\, e_1 = v^{j-1}\left(e_1^T \mathbf{M}\, e_1\right)$ for $1 \leq j \leq p$. Thus, $\mathbf{M}_{11} = 1/\hat{\Phi}(v)$.

Similarly, diagonal elements of the submatrix $\mathbf{A}_\Theta$, can be found as the previous case. The only modifications are that p and ϕ_i's are replaced by q and θ_i's. Therefore, if a matrix $\mathbf{N}$ is denoted as $[\mathbf{I} - v\mathbf{A}_\Theta]^{-1}$, the 1st and ith diagonal elements of the matrix are

$$N_{11} = \frac{1}{\hat{\Theta}(v)} \quad \text{and}$$

$$N_{ii} = \frac{1 - \sum\limits_{j=1}^{i-1} \hat{\theta}_j v^j}{1 - \sum\limits_{j=1}^{q} \hat{\theta}_j v^j} = \frac{1 - \sum\limits_{j=1}^{i-1} \hat{\theta}_j v^j}{\hat{\Theta}(v)}$$

where $2 \leq i \leq q$.

VITA

Name: Hyun Cheol Lee

Permanent Address: Hyundai APT. #106−1203 Woosan−Dong Buk−Gu

Gwang−Ju, Korea, 500−080

Educational Background: B.S. 1996. Department of Industrial Engineering

Korea University

M.S. 1999. Department of Industrial Engineering

Texas A&M University