Kalman filtering and statistical process control

Stephen V. Crowder
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
Crowder, Stephen V., "Kalman filtering and statistical process control " (1986). Retrospective Theses and Dissertations. 7991.
https://lib.dr.iastate.edu/rtd/7991
INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.
Crowder, Stephen V.

KALMAN FILTERING AND STATISTICAL PROCESS CONTROL

Iowa State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106
Kalman filtering and statistical process control

by

Stephen V. Crowder

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1986
# TABLE OF CONTENTS

## I. INTRODUCTION  
1. Background and the Kalman Model  
2. The Kalman Filter  
3. Applications of the Kalman Filter to Process Control

## II. KALMAN FILTERING AND STATISTICAL PROCESS CONTROL  
2. Kalman Filtering for a Process Mean  
3. Average Run Lengths for a GMA Chart for Sample Means  
4. Choice of Parameters for a GMA Chart for a Process Mean  
5. A Generalized Kalman Filter for a Process Variance  
6. Average Run Lengths for a GMA Chart for Sample Variances  
7. Choice of Parameters for a GMA Chart for a Process Variance

## III. ADAPTIVE KALMAN FILTERING AND STATISTICAL PROCESS CONTROL  
1. A Generalization of the Steady Model and Estimation of Variance Components  
2. An Adaptive Kalman Filtering Algorithm for a Process Mean  
3. Control Charting Alternatives for a Process Mean  
## IV. A UNIVARIATE PROCESS CONTROL MODEL

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. The General Linear Stochastic Control Model</td>
<td>97</td>
</tr>
<tr>
<td>B. A Special Case of the General Linear Stochastic Control Problem and an Alternative Performance Criterion</td>
<td>98</td>
</tr>
<tr>
<td>C. A Functional Equation</td>
<td>105</td>
</tr>
<tr>
<td>D. Solution of the Functional Equation</td>
<td>107</td>
</tr>
<tr>
<td>E. Complete State Information</td>
<td>113</td>
</tr>
<tr>
<td>F. Convergence Results</td>
<td>114</td>
</tr>
<tr>
<td>G. Asymptotic Average Risk - Optimal Control Policy</td>
<td>117</td>
</tr>
<tr>
<td>H. Asymptotic Average Risk - Shewhart Control Policy</td>
<td>119</td>
</tr>
<tr>
<td>I. Evaluation of Steady State Control Action Limits</td>
<td>127</td>
</tr>
</tbody>
</table>

## V. BIBLIOGRAPHY

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
</tr>
</tbody>
</table>

## VI. ACKNOWLEDGEMENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>133</td>
</tr>
</tbody>
</table>

## VII. APPENDIX

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Convergence of $V(\theta_t</td>
<td>y^t)$</td>
</tr>
<tr>
<td>B. Convergence of $\alpha_t$ Sequence</td>
<td>136</td>
</tr>
<tr>
<td>C. A Monotonicity Result</td>
<td>137</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

A. Background and the Kalman Model

The Kalman filter originally appeared in the engineering literature, in Kalman (1960) and Kalman and Bucy (1961). Since its introduction, it has been commonly used by control engineers and other physical scientists in such areas as missile trajectory and satellite orbit estimation.

The Kalman filter has also appeared extensively in the econometrics literature. Harrison (1967) uses a special case of the model as a tool in short-term sales forecasting. Sarris (1973) addresses the problem of estimating time-varying regression coefficients from a Bayesian point of view, and Sant (1977) applies generalized least squares to the same problem.

Until recently, however, the Kalman filter had not appeared in the statistical literature. Harrison and Stevens (1971, 1976) derived the Kalman filter from a Bayesian forecasting point of view. Duncan and Horn (1972) demonstrated the equivalence of Kalman filter theory and random parameter linear regression theory. While the original development of the Kalman filter is in a language foreign to statisticians, papers such as these have illustrated the filter's relation to linear models of regression and time series analysis. And because of its usefulness in applications, the Kalman filter is being viewed with increasing interest by statisticians.

Morrison and Pike (1977) applied the Kalman model to derive a short-term forecasting algorithm with time-varying parameters. Ledolter
(1979) used Kalman filtering techniques to calculate recursive estimates in regression and autoregressive integrated moving average (ARIMA) time series models. Sallas and Harville (1981) extended the Kalman model to derive recursive estimation equations for mixed models.

The derivation of the Kalman filter given here will follow that of Harrison and Stevens (1971).

The Kalman model can be described by the matrix equations

\[ Y_t = F_t \theta_t + \varepsilon_t \quad t = 1, 2, \ldots \]  
\[ \theta_t = G_t \theta_{t-1} + \iota_t \]

where \( F_t \) and \( G_t \) are known. The data at time \( t \) are represented by \( Y_t \), which may be either scalar or vector valued. The dependence of \( Y_t \) on \( \theta_t \), the unobservable state of nature, is described in (1.1), the observation equation. The observation error \( \varepsilon_t \) is assumed to be normally distributed with mean zero and known variance \( V_t \). The state of nature \( \theta_t \) is assumed to change with time according to (1.2), the system equation. The matrix \( G_t \) describes the transition of the state of nature from time \( t-1 \) to \( t \). The system equation error \( \iota_t \) is normally distributed with mean zero and known variance \( W_t \). One also assumes that \( \varepsilon_t \) and \( \iota_t \) are independent. Note that the matrices \( F_t \) and \( G_t \) of (1.1) and (1.2) may change with time, as also may the matrices \( V_t \) and \( W_t \).
B. The Kalman Filter

The Kalman filter is a recursive procedure for inference concerning the state of nature $\theta_t$ in (1.2). The derivation given here uses a Bayesian approach to estimating $\theta_t$ given the data through time $t$. Represent the data available through time $t$ as

$$Y^t = (Y_1, Y_2, \ldots, Y_t).$$

At time $t-1$ one's knowledge of $\theta_{t-1}$ can be expressed in terms of a posterior distribution for $\theta_{t-1}$ given $Y^{t-1}$,

$$\theta_{t-1}|Y^{t-1} \sim N(\hat{\theta}_{t-1}, Q_{t-1})$$

(1.3)

where $\hat{\theta}_{t-1}$ and $Q_{t-1}$ are the mean and variance of $\theta_{t-1}|Y^{t-1}$. Prior to observing $Y_t$, from (1.2) our knowledge of $\theta_t$ can be described by the conditional distribution

$$\theta_t|Y^{t-1} \sim N(G_t G_{t-1} \hat{\theta}_{t-1}, R_t)$$

(1.4)

where $R_t = G_t Q_{t-1} G_t' + W_t$. Now, let $e_t$ denote the error in predicting $Y_t$ given data through time $t-1$,

$$e_t = Y_t - \hat{Y}_t = Y_t - F_t G_t \hat{\theta}_{t-1}.$$

(1.5)
Since \( F_t, G_t, \) and \( \hat{\theta}_{t-1} \) are all known, observing \( Y_t \) is equivalent to observing \( e_t \). Since \( e_t \sim N(0, V_t) \) we have from (1.1) that

\[
e_t = Y_t - F_t \hat{G}_t \hat{\theta}_{t-1}
\]

\[
= F_t (\theta_t - G_t \hat{\theta}_{t-1}) + \epsilon_t
\]

and

\[
e_t | \theta_t, Y^{t-1} \sim N(F_t (\theta_t - G_t \hat{\theta}_{t-1}), V_t).
\]

(1.6)

Using a standard result from Anderson (1971) and the distributions for \( \theta_t | Y^{t-1} \) and \( e_t | \theta_t, Y^{t-1} \) we have that

\[
\begin{pmatrix}
\theta_t \\
e_t
\end{pmatrix} | Y^{t-1} \sim N\left(\begin{pmatrix}
G_t \hat{\theta}_{t-1} \\
F_t R_t V_t + F_t R_t F_t
\end{pmatrix}, \begin{pmatrix}
R_t & R_t F_t' \\
F_t R_t & V_t + F_t R_t F_t
\end{pmatrix}\right).
\]

(1.7)

Now, using (1.7) and conditioning on \( e_t \), the distribution of \( \theta_t | Y^t \) is equivalent to the distribution of \( \theta_t | e_t, Y^{t-1} \), namely

\[
\theta_t | Y^t \sim N(\hat{\theta}_t, Q_t),
\]

(1.8)

where

\[
\hat{\theta}_t = G_t \hat{\theta}_{t-1} + R_t F_t' (V_t + F_t R_t F_t)^{-1} e_t
\]

(1.9)

\[
Q_t = R_t - R_t F_t' (V_t + F_t R_t F_t)^{-1} F_t R_t
\]
and \[ e_t = Y_t - F_t \hat{\theta}_{t-1} \],

again, using standard results of multivariate analysis. So the cycle from (1.3) to (1.8) is completed, and the procedure moves on to time \( t+1 \). The recursive procedure described by (1.9) is begun at time zero by choosing \( \hat{\theta}_0 \) and \( \Theta_0 \) to be best guesses about the mean and the variance of \( \theta_0 \), respectively.

Note in (1.9) that the posterior mean of \( \theta_t | Y^t \) is the sum of two quantities, \( G_t \hat{\theta}_{t-1} \), and a multiple of the one step ahead forecast error \( e_t \). \( G_t \hat{\theta}_{t-1} \) is the mean of the prior distribution of \( \theta_t | Y^{t-1} \), and \( e_t \) is the error in predicting \( Y_t \) given the data through time \( t-1 \). Therefore, we can think of the Kalman filter as an updating procedure which combines a best prior guess about \( \theta_t \) with a correction factor depending on \( e_t \), a measure of how useful that prior guess has been in predicting the current observation.

Here we have assumed normality for the error vectors \( \epsilon_t \) and \( \nu_t \). The estimator \( \hat{\theta}_t \) in (1.9), which is the Kalman filter estimate, is the posterior mean for \( \theta_t \) given \( Y^t \) and is thus the Bayes estimator under squared error loss. Duncan and Horn (1972) show that even if the error vectors are not normally distributed, the Kalman filter estimator will still be the minimum mean square linear estimator provided the \( \epsilon_t \) and \( \nu_t \) are independent vectors with means zero and respective variances \( V_t \) and \( W_t \).
Harrison and Stevens (1976) present several examples illustrating that the Bayesian approach to forecasting includes many conventional methods such as linear regression, exponential smoothing, and linear time series models as special cases. One of the advantages of the Kalman filter approach to estimation and forecasting is the parametric structure of the model. At any given time, probabilistic information on the parameters is available in the form of a posterior distribution given all available data. Also, the sequential model definition (1.1, 1.2) describes how the parameters change in time, both systematically and as a result of random shocks.

The recursive nature of the algorithm (1.9) is important. In estimating and forecasting practice it means that the current posterior distribution of $\theta_t$ may be calculated from the most recent observation $y_t$, the posterior distribution of $\theta_{t-1|Y^t-1}$ and the current observation and system variances. Thus, it is not necessary to maintain the entire process history, increasing the Kalman filter's ease of use and computational efficiency.

C. Applications of the Kalman Filter to Process Control

The use of the Kalman filter as an estimation technique in statistical process control is increasing. In cases where prior information about the process is available, procedures based on the Kalman filter can be superior to the classical procedures like Shewhart and CUSUM control charts. Pike, Morrison, and Downing (1978) apply the
Kalman filter to special nuclear materials control and accountability and show it to be superior to the classical approaches. Phadke (1981) uses the Kalman filter as a tool in quality auditing. MacGregor (1973) and MacGregor and Wong (1980) use the Kalman filter to develop stochastic control theory for process application.

The objective in this thesis is to develop further applications of the Kalman filter as a tool in univariate process monitoring and optimal control. In Chapter II, we show that particular Kalman filter models for a process mean and variance lead to estimation of process parameters by geometric moving averages (GMA). Properties of the GMA as a process monitoring tool are studied using integral equations for moments of run length distributions. Extensions of the models in Chapter II are presented in Chapter III, along with adaptive Kalman filtering algorithms. Properties of the adaptive algorithms are then studied via simulation. In Chapter IV, the Kalman filter is used as a tool in stochastic control theory. We introduce a nonstandard but intuitively appealing cost structure for the optimal control problem. This structure leads to optimal policies that (unlike usual strategies for the problem) are consistent with Shewhart statistical process control philosophy. That is, process adjustment is called for only when evidence of misadjustment is strong. We also provide tables that allow implementation of these new policies.
II. KALMAN FILTERING AND STATISTICAL PROCESS CONTROL

A. Classical Evaluation of Control Procedures by Examination of Run Length Properties Under I.I.D. Models

In this chapter, we will use a special case of the Kalman filter model (1.1, 1.2) to develop an algorithm for monitoring a univariate process mean. A generalization of the Kalman filter model to the case of nonnormal observations will be used to develop an algorithm for monitoring a process variance. The properties of these algorithms will then be compared to the commonly used Shewhart-type monitoring of a process mean and variance not under the Kalman models, but rather under the classical "stable process" i.i.d. models.

Duncan (1974) gives a nice discussion of Shewhart-type variables control charts, including the X-bar chart, range chart, and s² chart. Using the Shewhart charting approach, a statistic $Q_t$, say, is computed from data collected at time $t$. If the $Q_t$ value is larger than a specified upper control limit $U$, say, or smaller than a lower control limit $L$, say, then the process is deemed "out of control." In this case, an investigation is undertaken to find the cause of the extreme value.

For the X-bar chart, with $\bar{X}_t$ the sample mean of $n$ observations at time $t$,

$$Q_t = \bar{X}_t$$

$$L = \text{Target} - 3 \cdot \sigma$$

$$\bar{X}_t$$
\[ U = \text{Target} + 3 \cdot \frac{\sigma}{\bar{X}_t} \]

where \( \sigma \) is the standard deviation of \( \bar{X}_t \). This Shewhart procedure is often justified using an assumption that the \( X_i \)'s are i.i.d. \( N(\mu, \sigma^2) \).

One way to quantify what any control procedure will do is to find the mean number of samples required to get an out of control signal assuming that the \( Q_t \)'s are i.i.d. with some known distribution. If \( T \) is the number of samples before a \( Q_t \) value first plots out of control, then the average run length (ARL) is the mean of \( T \) under an i.i.d. model. A vast literature describing Shewhart-type charts and their properties exists. See for example, Roberts (1966) and Duncan (1974).

### B. Kalman Filtering for a Process Mean

The model which we will use in the derivation of a procedure for monitoring a process mean is a generalization of the so-called steady model. The scalar steady model is described by

\[ y_t = \theta_t + \varepsilon_t \quad (2.1) \]

\[ \theta_t = \theta_{t-1} + \nu_t \quad t = 1, 2, \ldots \]

where \( \theta_t \) is the true, unknown state process mean at time \( t \), \( \{y_t, t = 1, 2, \ldots\} \) are observed outputs, and \( \{\varepsilon_t, t = 1, 2, \ldots\} \) and \( \{\nu_t, t = 1, 2, \ldots\} \) are independent sequences of independent normal random
variables with zero means and variances

\[ \text{Var}(\varepsilon_t) = \sigma^2_{\varepsilon} \]  
(2.2)

\[ \text{Var}(v_t) = \sigma^2_v \quad \text{for every } t. \]

Note that for the steady model, the parameter sequence \( \theta_t \) is a random walk. This model is appropriate for a situation where the most important characteristic of the process in question is its current true mean level, with persistent growth or decline being either absent or unimportant. Harrison and Stevens (1976) discuss this model as a special case of their Dynamic Linear Model. Harrison (1967) uses the model to describe customer demand for a steady selling product, and as a tool in short-term sales forecasting. Meinhold and Singpurwalla (1983) examine this model as an example in their overview of the Kalman filter model.

Also, the steady model has the same autocorrelation structure as a class of ARIMA(0, 1, 1) models of Box and Jenkins (1970), models which are widely used in practice. Box and Jenkins mention that models of this kind have often been found useful in inventory control problems, in representing some kinds of disturbance occurring in industrial processes, and in econometrics.

Note that if the steady model (2.1, 2.2) has \( v_t \equiv 0 \), an i.i.d. model results. For \( v_t \sim N(0, \sigma^2_v) \) \( t = 1, 2, \ldots \), the model allows for a random drift in the process mean. LaMotte and McWhorter (1978) have developed
an exact test for the presence of a random walk as specified by the steady model, while Nyblom (1983) compares several such tests. Also, diagnostic tools of time series analysis can be used to test the adequacy of the ARIMA(0, 1, 1) model representation.

Consider now the apparent slight generalization of (2.1)

\[ y_{it} = \theta_t + \epsilon_{it} \quad i = 1, 2, \ldots, n \quad (2.3) \]

\[ \theta_t = \theta_{t-1} + \nu_t \quad t = 1, 2, \ldots \]

where \( \theta_t \) is the true, unknown process mean at time \( t \), \( \{y_{it}, i = 1, 2, \ldots, n\} \) are observed outputs at time \( t \), each with mean \( \theta_t \), and \( \{\epsilon_{it}, i = 1, 2, \ldots, n; t = 1, 2, \ldots\} \) and \( \{\nu_t, t = 1, 2, \ldots\} \) are independent sequences of independent normal random variables with zero means and variances

\[ \text{Var}(\epsilon_{it}) = \sigma^2_\epsilon \quad i = 1, 2, \ldots, n \quad (2.4) \]

\[ \text{Var}(\nu_t) = \sigma^2_\nu \quad \text{for every } t. \]

This is of course a special case of the Kalman model (1.1, 1.2) with

\[ F_t \equiv 1_{-n} \]

\[ V_t \equiv \sigma^2_\epsilon \cdot I_n \]
and \( W_t \equiv \sigma_v^2 \),

where \( \mathbf{1}_n \) is an \( n \)-vector of ones and \( \mathbf{I}_n \) is the \( nxn \) identity matrix.

Let \( y_t \) be the vector of outputs observed at time \( t \) and \( y^t \) stand for \( (y_1, y_2, \ldots, y_t) \), the vector of observed outputs through time \( t \).

Our knowledge about \( \theta_{t-1} \) at time \( t-1 \) is represented by the posterior distribution of \( \theta_{t-1} \) given \( y^{t-1} \):

\[
\theta_{t-1} | y^{t-1} \sim N(\hat{\theta}_{t-1}, q_{t-1}) .
\]

Following the development in chapter one, let \( e_t \) be the error in predicting \( y_t \) from the point \( t-1 \), then

\[
e_t = y_t - \frac{1}{n} \theta_{t-1} = \frac{1}{n} (\theta_t - \hat{\theta}_{t-1}) + e_t
\]

where \( e_t \) is the vector of residuals at time \( t \).

Then,

\[
\begin{pmatrix}
\theta_t \\
e_t
\end{pmatrix}
| y^{t-1} \sim MVN \left( \begin{pmatrix}
\hat{\theta}_{t-1} \\
e_t
\end{pmatrix}
, \begin{pmatrix}
q_{t-1}+\sigma_v^2 & (q_{t-1}+\sigma_v^2)1' \\
(q_{t-1}+\sigma_v^2)1 & \sigma_v^2 I + (q_{t-1}+\sigma_v^2)11'
\end{pmatrix}
\right).
\]

Now, since \( \hat{\theta}_{t-1} \) is known at time \( t \), observing \( y_t \) is equivalent to observing \( e_t \). So the distribution of \( \theta_t | (y_t, y^{t-1}) \) is the same as that of
\( \theta_t \mid (e_t, y^{t-1}) \). From Anderson (1971),

\[
\theta_t \mid (e_t, y^{t-1}) \sim N(\hat{\theta}_t, q_t),
\]

where

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + (q_{t-1} + \sigma^2)\frac{1}{\sigma^2 I_n + (q_{t-1} + \sigma^2) I'_{11'}}^{-1} e_t
\]

and

\[
q_t = (q_{t-1} + \sigma^2) - (q_{t-1} + \sigma^2)\frac{1}{\sigma^2 I_n + (q_{t-1} + \sigma^2) I'_{11'}}^{-1} (q_{t-1} + \sigma^2)_{11'}
\]

These expressions reduce to

\[
\hat{\theta}_t = \hat{\theta}_{t-1} + \left( \frac{q_{t-1} + \sigma^2}{q_{t-1} + \sigma^2 + \sigma^2/n} \right) (\bar{y}_t - \hat{\theta}_{t-1})
\]

and

\[
q_t = (q_{t-1} + \sigma^2) \left[ \frac{\sigma^2/n}{\sigma^2/n + \sigma^2 + q_{t-1}} \right],
\]

where \( \bar{y}_t \) is the sample mean of outputs at time \( t \). Thus, the desired posterior distribution for \( \theta_t \) after time \( t \) is

\( \theta_t \mid y^t \sim N(\hat{\theta}_t, q_t) \).
The Kalman filter recursive estimates can then be determined from the equations

$$\hat{\theta}_t = (1 - k_t) \hat{\theta}_{t-1} + k_t \bar{y}_t,$$  \hspace{1cm} (2.5)

where

$$k_t = \left( \frac{q_{t-1} + \sigma_y^2}{q_{t-1} + \sigma_y^2 + \sigma_\varepsilon^2/n} \right)$$

and $q_t$ is given above.

Note that since $\sigma_y^2$ and $\sigma_\varepsilon^2$ are known, the posterior mean $\hat{\theta}_t$ depends on the data only through the sample means, and is an exponentially weighted moving average of $\bar{y}_1$, ..., $\bar{y}_t$. And for purposes of estimating $\theta_t$'s, the model (2.3, 2.4) is in fact equivalent to the steady model (2.1, 2.2) replacing $y_t$ with $\bar{y}_t$ and $\sigma_\varepsilon^2$ with $\sigma_\varepsilon^2/n$.

The sequence of posterior variances does not depend on the data. Appendix A shows that, in general, the sequence $q_0$, $q_1$, ..., converges rapidly to a limit.

Let $\Delta = \frac{\sigma_\varepsilon^2/n}{\sigma_y^2}$. Then using Appendix A,

$$q \equiv \lim_{t \to \infty} q_t = ((\Delta + \frac{1}{4})^{1/2} + \frac{1}{2})^{-1}(\sigma_\varepsilon^2/n).$$  \hspace{1cm} (2.6)

The sequence of constants $k_t$, known as the Kalman weights, also converge.

Using (2.5) and (2.6),
If the system variance $\sigma^2_v$ is large relative to the observation variance $\sigma^2_c/n$, $\Delta$ is near zero and $k$ is near one. In this case, the Kalman filter estimate places most of its weight on the current data $\bar{y}_t$. Conversely, if $\sigma^2_v$ is small relative to $\sigma^2_c/n$, $k$ is near zero and very little weight is placed on the current data.

Since the convergence of $q_t$ and $k_t$ is rapid, we will use their steady state values, and the recursive estimation scheme reduces to

$$\hat{\theta}_t = (1 - k) \hat{\theta}_{t-1} + ky_t,$$

for

$$k = ((\Delta + \frac{1}{4})^{1/2} + \frac{1}{2})^{-1},$$

with

$$\Delta = \frac{\sigma^2_c}{\sigma^2_v}.$$

Given some starting value $\hat{\theta}_0$, our estimate of $\theta_t$ after time $t$ is a geometric moving average of $\hat{\theta}_0$ and $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_t$.

From the Kalman filtering theory, we know that if model (2.3, 2.4) is correct, the best estimate in terms of minimum mean squared error of $\theta_t$ based on the data through time $t$ is the Kalman filter estimate. Here we want to compare ARL properties of a Kalman filter (geometric moving average) control chart to those of a usual Shewhart X-bar chart, under
a classical i.i.d. model for the $\overline{y}_t$'s. We will also indicate how the run length distribution can be studied in detail, by numerical evaluation of its higher moments.

C. Average Run Lengths for a GMA Chart for Sample Means

Suppose we are plotting a geometric moving average (GMA) of sample statistics $y_t$. (For early discussions of the GMA chart, see Roberts (1959) or Wetherill (1977).) The successive values plotted can be described by

$$Q_t = \alpha Q_{t-1} + (1 - \alpha)y_t \quad 0 < \alpha < 1, \quad t = 1, 2, \ldots . \quad (2.7)$$

Here we wish to study the ARL properties of a procedure based on the $Q_t$'s when the $y_t$'s are assumed to be i.i.d. $N(\mu, \sigma^2)$. Using the $Q_t$'s to monitor the process mean level, the process will be deemed "out of control" if $Q_t$ is too large or too small. Without loss of generality, we will take the desired (target) mean value to be $\mu = 0$, and the known process variance to be $\sigma^2 = 1$. At time $t$, we will conclude there is a lack of control if

$$|Q_t| > h, \text{ for } h \text{ a specified constant.}$$

Let $L(u)$ be the ARL given that the GMA starts at $Q_0 = u$. Using (2.7),
\[ L(u) = 1 \cdot P(|\alpha u + (1-\alpha)y| > h) \]

\[ + \int (1 + L(\alpha u + (1-\alpha)y)) f_\mu(y) \, dy \right|_{|\alpha u + (1-\alpha)y| < h} \]

where \( f_\mu(y) \) is the \( N(\mu, 1) \) density function of \( y \). So

\[ L(u) = 1 + \int L(\alpha u + (1-\alpha)y) f_\mu(y) \, dy \right|_{|\alpha u + (1-\alpha)y| < h} \]

\[ = 1 + \int L(\alpha u + (1-\alpha)y) f_\mu(y) \, dy \right|_{(h-\alpha u)/(1-\alpha)} \]

\[ = 1 + \frac{1}{(1-\alpha)} \int_{-h}^{h} L(t) f \left( \frac{t-\alpha u}{1-\alpha} \right) dt \quad . \tag{2.8} \]

This integral equation for \( L(\cdot) \) is a Fredholm equation of the second kind. For a complete discussion of methods of approximating the unknown \( L(\cdot) \), see C. T. H. Baker (1977). The simplest, and most readily applied in practice is the quadrature method, which will be briefly described here.

The general Fredholm equation of the second kind is
\[ f(x) = g(x) + \lambda \int_{a}^{b} k(x,y)f(y)dy , \]  

(2.9)

where \( f(x) \) is unknown, \( a \) and \( b \) are finite, \( k(x,y) \) (the kernel) and \( g(x) \) are continuous known functions, and \( \lambda \) is a known constant. Given a quadrature rule of the form \( \sum_{j=0}^{n} w_j \phi(a_j) \) for approximating an integral \( \int_{a}^{b} \phi(y)dy \), equation (2.9) can be replaced by

\[ \tilde{f}(x) = g(x) + \lambda \sum_{j=0}^{n} w_j k(x,a_j) \tilde{f}(a_j) , \]  

(2.10)

in which \( \tilde{f}(x) \) may be regarded as an approximation to \( f(x) \). Equation (2.10) is a functional equation whose solution may be found by setting \( x = a_i \) (\( i = 0, 1, 2, \ldots, n \)) in (2.10) to obtain the equations

\[ \tilde{f}(a_i) = g(a_i) + \lambda \sum_{j=0}^{n} w_j k(a_i,a_j) \tilde{f}(a_j) \quad i = 0, 1, \ldots, n. \]

If values \( \tilde{f}(a_0), \tilde{f}(a_1), \ldots, \tilde{f}(a_n) \) satisfying these equations can be found, we can then obtain the solution of equation (2.10) for all \( x \in [a,b] \) by setting

\[ \tilde{f}(x) = g(x) + \lambda \sum_{j=0}^{n} w_j k(x,a_j) \tilde{f}(a_j) . \]  

(2.11)

Equation (2.11) gives the so-called Nyström approximation of \( f(x) \).
A Gaussian quadrature rule was used as described above to approximate \( L(u) \) in (2.8). Values of \( L(0) \), the ARL for the GMA control scheme starting at zero, were approximated numerically and appear in Table (2.1). On the GMA chart we plot values of

\[
Q_t = \alpha Q_{t-1} + (1-\alpha)\bar{y}_t \quad 0 < \alpha < 1.
\]

And for large \( t \), under an i.i.d. model, the variance of \( Q_t \) is approximately

\[
\sigma_Q^2 \equiv \text{Var}(Q_t) = \left( \frac{1-\alpha}{1+\alpha} \right) \frac{\sigma^2}{n}, \quad (2.12)
\]

where the \( \bar{y}_t \)'s are \( N(\mu, \frac{\sigma^2}{n}) \). We will specify control limits for our GMA chart in terms of \( \sigma_Q \). That is, the limits will be taken to be \( \pm L\sigma_Q \).

Thus, a particular \( L(0) \) value in Table 2.1 depends on the true mean \( \mu \), the weighting constant \( \alpha \), and the constant \( L \). In Table 2.1 the true mean is expressed in units of \( \sigma/\sqrt{n} \), the standard deviation of \( \bar{y}_t \). For example, the ARL given that the GMA starts at \( Q_0 = 0 \), for \((\sqrt{n}\mu/\sigma, \alpha, L) = (.25, .50, 2.50) \) is approximately 58.3. The values in Table 2.1 are consistent with the simulation results given by Roberts (1959), and the numerical approximations obtained by Robinson and Ho (1978) using a technique entirely different from ours. These ARL values provide a way to compare the GMA chart to the usual Shewhart X-bar chart. Also, these values can be used to choose appropriately \( \alpha \) and \( L \), the parameters of the GMA chart.
Table 2.1. Average run lengths for two-sided geometric moving average charts

<table>
<thead>
<tr>
<th>( \sqrt{n}/\sigma )</th>
<th>L=0.0</th>
<th>L=0.25</th>
<th>L=0.50</th>
<th>L=0.75</th>
<th>L=0.90</th>
<th>L=2.0</th>
<th>L=2.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>21.98</td>
<td>22.88</td>
<td>26.45</td>
<td>38.56</td>
<td>73.28</td>
<td>40.90</td>
<td>42.25</td>
</tr>
<tr>
<td>0.25</td>
<td>19.13</td>
<td>18.86</td>
<td>20.12</td>
<td>24.83</td>
<td>34.49</td>
<td>34.53</td>
<td>33.07</td>
</tr>
<tr>
<td>0.50</td>
<td>13.70</td>
<td>12.34</td>
<td>11.89</td>
<td>12.74</td>
<td>15.53</td>
<td>23.23</td>
<td>19.78</td>
</tr>
<tr>
<td>0.75</td>
<td>9.21</td>
<td>7.86</td>
<td>7.29</td>
<td>7.62</td>
<td>9.36</td>
<td>14.67</td>
<td>11.64</td>
</tr>
<tr>
<td>1.00</td>
<td>6.25</td>
<td>5.26</td>
<td>4.91</td>
<td>5.24</td>
<td>6.62</td>
<td>9.41</td>
<td>7.31</td>
</tr>
<tr>
<td>1.25</td>
<td>4.40</td>
<td>3.76</td>
<td>3.59</td>
<td>3.95</td>
<td>5.13</td>
<td>6.29</td>
<td>4.95</td>
</tr>
<tr>
<td>1.50</td>
<td>3.24</td>
<td>2.84</td>
<td>2.80</td>
<td>3.19</td>
<td>4.20</td>
<td>4.41</td>
<td>3.58</td>
</tr>
<tr>
<td>1.75</td>
<td>2.49</td>
<td>2.26</td>
<td>2.29</td>
<td>2.68</td>
<td>3.57</td>
<td>3.24</td>
<td>2.75</td>
</tr>
<tr>
<td>2.00</td>
<td>2.00</td>
<td>1.88</td>
<td>1.95</td>
<td>2.32</td>
<td>3.12</td>
<td>2.49</td>
<td>2.21</td>
</tr>
<tr>
<td>2.25</td>
<td>1.67</td>
<td>1.61</td>
<td>1.70</td>
<td>2.06</td>
<td>2.78</td>
<td>2.00</td>
<td>1.85</td>
</tr>
<tr>
<td>2.50</td>
<td>1.45</td>
<td>1.42</td>
<td>1.51</td>
<td>1.85</td>
<td>2.52</td>
<td>1.67</td>
<td>1.60</td>
</tr>
<tr>
<td>2.75</td>
<td>1.29</td>
<td>1.29</td>
<td>1.37</td>
<td>1.69</td>
<td>2.32</td>
<td>1.45</td>
<td>1.42</td>
</tr>
<tr>
<td>3.00</td>
<td>1.19</td>
<td>1.19</td>
<td>1.26</td>
<td>1.55</td>
<td>2.16</td>
<td>1.29</td>
<td>1.29</td>
</tr>
<tr>
<td>3.25</td>
<td>1.12</td>
<td>1.13</td>
<td>1.18</td>
<td>1.43</td>
<td>2.03</td>
<td>1.19</td>
<td>1.19</td>
</tr>
<tr>
<td>3.50</td>
<td>1.07</td>
<td>1.08</td>
<td>1.12</td>
<td>1.32</td>
<td>1.93</td>
<td>1.12</td>
<td>1.13</td>
</tr>
<tr>
<td>3.75</td>
<td>1.04</td>
<td>1.05</td>
<td>1.08</td>
<td>1.24</td>
<td>1.83</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td>4.00</td>
<td>1.02</td>
<td>1.03</td>
<td>1.05</td>
<td>1.17</td>
<td>1.73</td>
<td>1.04</td>
<td>1.05</td>
</tr>
<tr>
<td>4.25</td>
<td>1.01</td>
<td>1.02</td>
<td>1.04</td>
<td>1.16</td>
<td>1.69</td>
<td>1.03</td>
<td>1.04</td>
</tr>
<tr>
<td>4.50</td>
<td>1.00</td>
<td>1.01</td>
<td>1.03</td>
<td>1.15</td>
<td>1.62</td>
<td>1.02</td>
<td>1.03</td>
</tr>
<tr>
<td>4.75</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
<td>1.14</td>
<td>1.58</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td>5.00</td>
<td>0.98</td>
<td>0.98</td>
<td>1.00</td>
<td>1.13</td>
<td>1.54</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>5.25</td>
<td>0.97</td>
<td>0.97</td>
<td>0.99</td>
<td>1.12</td>
<td>1.51</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>5.50</td>
<td>0.96</td>
<td>0.96</td>
<td>0.98</td>
<td>1.11</td>
<td>1.48</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>5.75</td>
<td>0.95</td>
<td>0.95</td>
<td>0.97</td>
<td>1.10</td>
<td>1.45</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>6.00</td>
<td>0.94</td>
<td>0.94</td>
<td>0.96</td>
<td>1.09</td>
<td>1.42</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>6.25</td>
<td>0.93</td>
<td>0.93</td>
<td>0.95</td>
<td>1.08</td>
<td>1.39</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>6.50</td>
<td>0.92</td>
<td>0.92</td>
<td>0.94</td>
<td>1.07</td>
<td>1.36</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>6.75</td>
<td>0.91</td>
<td>0.91</td>
<td>0.93</td>
<td>1.06</td>
<td>1.33</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>7.00</td>
<td>0.90</td>
<td>0.90</td>
<td>0.92</td>
<td>1.05</td>
<td>1.30</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>7.25</td>
<td>0.89</td>
<td>0.89</td>
<td>0.91</td>
<td>1.04</td>
<td>1.27</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>7.50</td>
<td>0.88</td>
<td>0.88</td>
<td>0.90</td>
<td>1.03</td>
<td>1.24</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>7.75</td>
<td>0.87</td>
<td>0.87</td>
<td>0.89</td>
<td>1.02</td>
<td>1.21</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>8.00</td>
<td>0.86</td>
<td>0.86</td>
<td>0.88</td>
<td>1.01</td>
<td>1.18</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>8.25</td>
<td>0.85</td>
<td>0.85</td>
<td>0.87</td>
<td>1.00</td>
<td>1.15</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>8.50</td>
<td>0.84</td>
<td>0.84</td>
<td>0.86</td>
<td>0.99</td>
<td>1.12</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>8.75</td>
<td>0.83</td>
<td>0.83</td>
<td>0.85</td>
<td>0.98</td>
<td>1.09</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>9.00</td>
<td>0.82</td>
<td>0.82</td>
<td>0.84</td>
<td>0.97</td>
<td>1.06</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>9.25</td>
<td>0.81</td>
<td>0.81</td>
<td>0.83</td>
<td>0.96</td>
<td>1.03</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>9.50</td>
<td>0.80</td>
<td>0.80</td>
<td>0.82</td>
<td>0.95</td>
<td>1.00</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>9.75</td>
<td>0.79</td>
<td>0.79</td>
<td>0.81</td>
<td>0.94</td>
<td>0.97</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>10.00</td>
<td>0.78</td>
<td>0.78</td>
<td>0.80</td>
<td>0.93</td>
<td>0.94</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>10.25</td>
<td>0.77</td>
<td>0.77</td>
<td>0.79</td>
<td>0.92</td>
<td>0.92</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>10.50</td>
<td>0.76</td>
<td>0.76</td>
<td>0.78</td>
<td>0.91</td>
<td>0.89</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>10.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.77</td>
<td>0.90</td>
<td>0.87</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>11.00</td>
<td>0.74</td>
<td>0.74</td>
<td>0.76</td>
<td>0.89</td>
<td>0.85</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>11.25</td>
<td>0.73</td>
<td>0.73</td>
<td>0.75</td>
<td>0.88</td>
<td>0.83</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>11.50</td>
<td>0.72</td>
<td>0.72</td>
<td>0.74</td>
<td>0.87</td>
<td>0.80</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>11.75</td>
<td>0.71</td>
<td>0.71</td>
<td>0.73</td>
<td>0.86</td>
<td>0.78</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>12.00</td>
<td>0.70</td>
<td>0.70</td>
<td>0.72</td>
<td>0.85</td>
<td>0.75</td>
<td>0.72</td>
<td>0.72</td>
</tr>
</tbody>
</table>
Table 2.1. (continued)

<table>
<thead>
<tr>
<th>$\sqrt{n}/\sigma$</th>
<th>$L=2.5$</th>
<th>$L=2.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha=0.0$</td>
<td>.25</td>
<td>.50</td>
</tr>
<tr>
<td>0.00</td>
<td>80.52</td>
<td>82.49</td>
</tr>
<tr>
<td>0.25</td>
<td>65.77</td>
<td>61.07</td>
</tr>
<tr>
<td>0.50</td>
<td>41.49</td>
<td>33.26</td>
</tr>
<tr>
<td>0.75</td>
<td>24.61</td>
<td>18.05</td>
</tr>
<tr>
<td>1.00</td>
<td>14.92</td>
<td>10.57</td>
</tr>
<tr>
<td>1.25</td>
<td>9.46</td>
<td>6.75</td>
</tr>
<tr>
<td>1.50</td>
<td>6.30</td>
<td>4.65</td>
</tr>
<tr>
<td>1.75</td>
<td>4.41</td>
<td>3.43</td>
</tr>
<tr>
<td>2.00</td>
<td>3.24</td>
<td>2.67</td>
</tr>
<tr>
<td>2.25</td>
<td>2.49</td>
<td>2.17</td>
</tr>
<tr>
<td>2.50</td>
<td>2.00</td>
<td>1.83</td>
</tr>
<tr>
<td>2.75</td>
<td>1.67</td>
<td>1.59</td>
</tr>
<tr>
<td>3.00</td>
<td>1.45</td>
<td>1.41</td>
</tr>
<tr>
<td>3.25</td>
<td>1.29</td>
<td>1.29</td>
</tr>
<tr>
<td>3.50</td>
<td>1.19</td>
<td>1.19</td>
</tr>
<tr>
<td>3.75</td>
<td>1.12</td>
<td>1.13</td>
</tr>
<tr>
<td>4.00</td>
<td>1.07</td>
<td>1.08</td>
</tr>
<tr>
<td>$\sqrt{n\mu/\sigma}$</td>
<td>$L=3.0$</td>
<td>$L=3.5$</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>$\alpha=0.0$</td>
<td>.25</td>
<td>.50</td>
</tr>
<tr>
<td>0.00</td>
<td>370.40</td>
<td>374.50</td>
</tr>
<tr>
<td>0.25</td>
<td>281.15</td>
<td>245.76</td>
</tr>
<tr>
<td>0.50</td>
<td>155.22</td>
<td>110.95</td>
</tr>
<tr>
<td>0.75</td>
<td>81.22</td>
<td>50.92</td>
</tr>
<tr>
<td>1.00</td>
<td>43.89</td>
<td>25.64</td>
</tr>
<tr>
<td>1.25</td>
<td>24.96</td>
<td>14.26</td>
</tr>
<tr>
<td>1.50</td>
<td>14.97</td>
<td>8.72</td>
</tr>
<tr>
<td>1.75</td>
<td>9.47</td>
<td>5.80</td>
</tr>
<tr>
<td>2.00</td>
<td>6.30</td>
<td>4.15</td>
</tr>
<tr>
<td>2.25</td>
<td>4.41</td>
<td>3.16</td>
</tr>
<tr>
<td>2.50</td>
<td>3.24</td>
<td>2.52</td>
</tr>
<tr>
<td>2.75</td>
<td>2.49</td>
<td>2.09</td>
</tr>
<tr>
<td>3.00</td>
<td>2.00</td>
<td>1.79</td>
</tr>
<tr>
<td>3.25</td>
<td>1.67</td>
<td>1.57</td>
</tr>
<tr>
<td>3.50</td>
<td>1.45</td>
<td>1.41</td>
</tr>
<tr>
<td>3.75</td>
<td>1.29</td>
<td>1.29</td>
</tr>
<tr>
<td>4.00</td>
<td>1.19</td>
<td>1.20</td>
</tr>
</tbody>
</table>
Table 2.1. (continued)

<table>
<thead>
<tr>
<th>$\sqrt{n \mu/\sigma}$</th>
<th>L=0.0</th>
<th>L=0.25</th>
<th>L=0.50</th>
<th>L=0.75</th>
<th>L=0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>15787.2</td>
<td>15806.3</td>
<td>16051.3</td>
<td>18069.9</td>
<td>26240.4</td>
</tr>
<tr>
<td>0.25</td>
<td>10090.2</td>
<td>7984.40</td>
<td>5576.65</td>
<td>2998.63</td>
<td>1387.89</td>
</tr>
<tr>
<td>0.50</td>
<td>4236.81</td>
<td>2520.47</td>
<td>1233.62</td>
<td>406.11</td>
<td>130.08</td>
</tr>
<tr>
<td>0.75</td>
<td>1729.98</td>
<td>833.01</td>
<td>324.04</td>
<td>90.06</td>
<td>37.73</td>
</tr>
<tr>
<td>1.00</td>
<td>740.64</td>
<td>303.84</td>
<td>104.22</td>
<td>31.65</td>
<td>19.44</td>
</tr>
<tr>
<td>1.25</td>
<td>335.59</td>
<td>122.96</td>
<td>41.15</td>
<td>15.81</td>
<td>12.90</td>
</tr>
<tr>
<td>1.50</td>
<td>161.04</td>
<td>55.31</td>
<td>19.74</td>
<td>9.90</td>
<td>9.67</td>
</tr>
<tr>
<td>1.75</td>
<td>81.80</td>
<td>27.67</td>
<td>11.23</td>
<td>7.11</td>
<td>7.77</td>
</tr>
<tr>
<td>2.00</td>
<td>43.96</td>
<td>15.36</td>
<td>7.31</td>
<td>5.55</td>
<td>6.51</td>
</tr>
<tr>
<td>2.25</td>
<td>24.96</td>
<td>9.40</td>
<td>5.27</td>
<td>4.57</td>
<td>5.63</td>
</tr>
<tr>
<td>2.50</td>
<td>14.97</td>
<td>6.27</td>
<td>4.08</td>
<td>3.90</td>
<td>4.96</td>
</tr>
<tr>
<td>2.75</td>
<td>9.47</td>
<td>4.51</td>
<td>3.34</td>
<td>3.42</td>
<td>4.45</td>
</tr>
<tr>
<td>3.00</td>
<td>6.30</td>
<td>3.45</td>
<td>2.84</td>
<td>3.05</td>
<td>4.05</td>
</tr>
<tr>
<td>3.25</td>
<td>4.41</td>
<td>2.76</td>
<td>2.48</td>
<td>2.77</td>
<td>3.72</td>
</tr>
<tr>
<td>3.50</td>
<td>3.24</td>
<td>2.30</td>
<td>2.22</td>
<td>2.54</td>
<td>3.44</td>
</tr>
<tr>
<td>3.75</td>
<td>2.49</td>
<td>1.97</td>
<td>2.01</td>
<td>2.36</td>
<td>3.22</td>
</tr>
<tr>
<td>4.00</td>
<td>2.00</td>
<td>1.73</td>
<td>1.84</td>
<td>2.21</td>
<td>3.03</td>
</tr>
</tbody>
</table>
To study the run length distribution of the GMA chart in greater detail, it is possible to use an integral equation more general than (2.8). We proceed to derive this equation.

Let \( y_t \) be a statistic observed at time \( t \), \( Q_t = \alpha Q_{t-1} + (1-\alpha)y_t \), \( a \) be the lower control limit on the GMA chart, and \( b \) be the upper control limit on the GMA chart. Here \( Q_t \) represents the value plotted at time \( t \), a GMA of the \( y_t \)'s that we assume are i.i.d. with density \( f(y) \). We are interested in the run length, or number of such values which will be plotted until one falls outside the interval \((a, b)\). Let \( N \) stand for this run length and let

\[
P(n, u) = P(N=n \text{ given that the GMA starts at } Q_0 = u).
\]

If \( u \) is the current plotted value and \( y \) is observed, then \( au + (1-\alpha)y \) is the next value plotted. If \( a \leq au + (1-\alpha)y \leq b \), the run continues, otherwise it is terminated. So for \( n \geq 2 \),

\[
P(n, u) = \int_{\{a \leq au + (1-\alpha)y \leq b\}} P(n-1, au + (1-\alpha)y)f(y)dy
\]

\[
= \frac{(b-au)/(1-\alpha)}{(a-au)/(1-\alpha)}
\]

\[
= \frac{1}{1-\alpha} \int_{a}^{b} P(n-1, y)f(\frac{y-au}{1-\alpha})dy. \tag{2.13}
\]
Now, for a function $g$, define

$$G_{\beta}(u) = \sum_{n=1}^{\infty} g(n+\beta)P(n,u). \quad (2.14)$$

$G_{\beta}(u)$ is clearly the mean of $g(N+\beta)$ given a start at $u$.

Then,

$$G_{\gamma}(u) = \sum_{n=1}^{\infty} g(n)P(n,u)$$

$$= g(1) \cdot P(1,u) + \sum_{n=2}^{\infty} g(n)P(n,u) .$$

Using (2.13),

$$G_{\gamma}(u) = g(1) \cdot P(1,u) + \frac{1}{1-\alpha} \int_a^b \sum_{n=2}^{\infty} g(n)P(n-1,y)f\left(\frac{y-\alpha u}{1-\alpha}\right)dy$$

$$= g(1) \cdot P(1,u) + \frac{1}{1-\alpha} \int_a^b G_{\gamma}(y)f\left(\frac{y-\alpha u}{1-\alpha}\right)dy .$$

Since $P(1,u) = 1 - P(a \leq \alpha u + (1-\alpha)y \leq b)$

$$= 1 - F\left(\frac{b-\alpha u}{1-\alpha}\right) + F\left(\frac{a-\alpha u}{1-\alpha}\right) ,$$

$$G_{\gamma}(u) = g(1) \cdot (1-F\left(\frac{b-\alpha u}{1-\alpha}\right) + F\left(\frac{a-\alpha u}{1-\alpha}\right)) + \frac{1}{1-\alpha} \int_a^b G_{\gamma}(y)f\left(\frac{y-\alpha u}{1-\alpha}\right)dy ,$$

where $F$ is the c.d.f. of $y$. \quad (2.15)
Equation (2.15) can be used to obtain integral equations for the mean of many useful functions of $N$. For example, taking $g(x) = x$ yields an expression for the mean run length, as in (2.8). Setting $g(x) = x^2$ produces an equation for the second moment of the run length. From (2.15) we then have that

$$V(u) = 1 + \frac{2}{1-\alpha} \int_a^b L(y)f\left(\frac{y-\alpha u}{1-\alpha}\right)dy + \frac{1}{1-\alpha} \int_a^b V(y)f\left(\frac{y-\alpha u}{1-\alpha}\right)dy,$$

where $L(u)$ and $V(u)$ are the first and second moments, respectively, of the run length given a start at $u$. Once we have an approximation for $L(u)$ from (2.8), we can use this expression to approximate $V(u)$. Setting $g(x) = e^{tx}$ yields an equation for the moment generating function of $N$. Again, using (2.15), we have that

$$e^{-t} \text{MGF}(t,u) = 1 - F\left(\frac{b-\alpha u}{1-\alpha}\right) + F\left(\frac{a-\alpha u}{1-\alpha}\right) + \frac{1}{1-\alpha} \int_a^b \text{MGF}(t,y)f\left(\frac{y-\alpha u}{1-\alpha}\right)dy,$$

where $\text{MGF}(t,u)$ is the moment generating function of the run length given a start at $u$. So it is possible to study the run length distribution in detail, by applying (2.15) and the numerical integration methods discussed earlier.
D. Choice of Parameters for a GMA Chart for a Process Mean

Various criteria for designing a quality control chart for monitoring a process mean have been suggested. Page (1961) recommends designing cumulative sum (CUSUM) charts to have specified ARL values at $\mu = 0$, the target mean, and at $\mu = \mu_\text{shift}$, where $\mu_\text{shift}$ is the smallest shift in the mean considered important enough to be detected quickly. Robinson and Ho (1978) make the same recommendation for constructing a GMA chart and outline a 3-step design procedure. Roberts (1966) compares the performance of several types of control charts, designed so that the ARL values are the same at $\mu = 0$, on the basis of their ARL values for nonzero $\mu$'s.

Yashchin (1985) proposes a nine step procedure for designing a CUSUM chart, using knowledge of the run length distribution (rather than just summarizations of it such as the ARL) to control the probability of an early false signal. Montgomery (1980) summarizes the economic design of control charts to minimize an expected total cost. He concludes that the economic design of control charts is used infrequently in practice because it is complex and input variables are usually not known precisely. When the required information is available, the economic design approach is preferred. When such information is not available, we recommend an approach due to Woodall (1985). Although his proposal is made for the design of CUSUM charts, the ideas apply equally well to the design of GMA charts, and will be described here.

Woodall's approach is to first specify a region of acceptable values for the process mean. A control chart is then designed to have
specified ARL values at two particular shifts in the underlying process mean. These shifts correspond to the largest shift $\mu_0 \geq 0$ considered to be of little or no practical importance, and the smallest shift $\mu_1 \geq \mu_0$ such that any shift larger than $\mu_1$ is considered important enough to be detected quickly. Clearly, we desire the ARL to be large whenever $|\mu| \leq \mu_0$ and small whenever $|\mu| \geq \mu_1$. Woodall uses the values $\mu_0$ and $\mu_1$ to define what he calls in-control, indifference, and out-of-control regions. The in-control region contains all values of $\mu$ such that $0 \leq |\mu| \leq \mu_0$ corresponding to the set of shifts considered to be of no practical importance. He argues that it may not be practical to make the fine adjustments to the process required to correct small deviations from the target $\mu = 0$ and that a $\mu_0 \neq 0$ can be appropriate. It can be the case that a small shift in the mean will not significantly increase the percentage of nonconforming product. Also, attempts to adjust a process when the shift is small can lead to over-correction and introduce extra variability into the process. And if some variability is tolerated in the quality characteristic, then it seems reasonable to tolerate some small shifts in the underlying mean.

The region of indifference contains all values of $\mu$ such that $\mu_0 \leq |\mu| \leq \mu_1$. The out-of-control region contains all values of $\mu$ such that $|\mu| > \mu_1$, the region for which a control procedure should give an out-of-control signal quickly. We will use the symbols $M_0$, I, and $M_1$ for the in-control, indifference, and out-of-control regions, respectively. That is,
\[ M_0 = \{ \mu | 0 \leq |\mu| \leq \mu_0 \}, \]
\[ I = \{ \mu | \mu_0 < |\mu| < \mu_1 \}, \]
and
\[ M_1 = \{ \mu | \mu_1 < |\mu| \}. \]

An appropriate control chart will be one with ARL of at least \( L_0 \) when \( \mu \in M_0 \) and at most \( L_1 \) when \( \mu \in M_1 \). To compare two control procedures meeting these minimal requirements, say A and B, one can compare their ARL profiles over the regions \( M_0 \) and \( M_1 \), but not over \( I \), the indifference region. If the ARL of procedure A is greater than that of B for \( \mu \in M_0 \) and less than that of B for \( \mu \in M_1 \), the procedure A is said to be uniformly better than procedure B. Clearly, it could also be the case that procedure A is preferable to B only over certain intervals of \( \mu \) values. Then, neither procedure A nor B is uniformly preferable. For example, when comparing a Shewhart chart to a CUSUM chart, it is typical that neither is uniformly preferable.

Using Table 2.1, two GMA procedures for normal observations can be compared, based on their respective ARL profiles. Our present goal is to design a GMA control chart to keep the true process mean \( \mu \) in the region \( M_0 \). In the present context, where the datum at time \( t \) consists of \( \bar{y}_t \), \( \sqrt{n} |\mu| / \sigma \) denotes the shift in the process mean measured in units of the standard deviation of the sample mean. On the GMA chart we plot values of

\[ Q_t = \alpha Q_{t-1} + (1-\alpha)\bar{y}_t \]
and specify control limits of the form ± L·σ Q, where σ Q is given by (2.12).

Designing an appropriate GMA chart consists of choosing (α, L, n) to satisfy Woodall's criteria. The following sequence of steps will lead to a GMA scheme which approximately satisfies the proposed criteria.

**Step 1.** Choose the regions M₀, I, M₁, and specify the two ARL values L₀ and L₁. (An economic analysis might be required to make this choice of μ₀ and μ₁.)

**Step 2.** For sample sizes n = 1, 2, 3, ... construct a table of values of (α, L) for which ARL = L₀ when μ = μ₀. These tables can be constructed from Table 2.1 by interpolation. Or the integral equation (2.8) for the ARL can be used to determine the values more precisely.

**Step 3.** Searching through the tables from Step 2, find the smallest value of n, n* say, for which the ARL at μ = μ₁ can approximate L₁.

**Step 4.** Given n*, if there is more than one (α, L) combination that has ARL = L₁ when μ = μ₁, take the one which appears to have the "best" ARL profile over M₀ and M₁, bearing in mind that there may not exist a uniformly best choice.
Step 5. Calculate the control limits $\pm L \cdot \sigma_Q$.

These steps are illustrated below, for the special case $\mu_0 = 0$. Table 2.2, used below, was taken from Robinson and Ho (1978).

Example 2.1

Suppose a two-sided GMA chart is to be designed for controlling a process such that the chart will yield an ARL of 500 when the process is exactly on target, and an ARL of 5 when the process mean has shifted one standard deviation. Here $\mu_0 = 0$, $L_0 = 500$, $L_1 = 5$, and $\mu_1/\sigma = 1$. Table 2.2, constructed according to Step 2, gives values of $(\alpha, L)$ for which $\text{ARL} = 500$ when $\mu = 0$. Looking down the first column of Table 2.2 headed $\sqrt{n}\mu/\sigma$, we locate $\sqrt{n}\mu/\sigma = 1.50$ with $(\alpha, L)$ combinations (.80, 2.972), (.70, 3.027), and (.60, 3.056) having ARL values 5.43, 5.44, and 5.79, respectively. Since $\mu_1/\sigma = 1$, we must have $\sqrt{n} > 1.50$ or $n > 2.25$ to obtain an ARL of 5, so take $n^* = 3$. Now looking down columns of Table 2.2 at values of $\sqrt{n}\mu/\sigma > \sqrt{3} \approx 1.73$, find the $(\alpha, L)$ combination with ARL $\approx 5$ for $\sqrt{n}\mu/\sigma \approx 1.73$ which has the "best" ARL profile for $\sqrt{n}\mu/\sigma > 1.73$, i.e., for $\mu$ in the out-of-control region. In this example, the appropriate choice of $(\alpha, L, n)$ appears to be approximately (.43, 3.076, 3), found by interpolation in Table 2.2. The control limits for the two-sided GMA then become $\pm L \cdot \sigma_Q$ where

$$L \sigma_Q = L \cdot \left(\frac{1-\alpha}{1+\alpha}\right)^{1/2} \frac{\sigma}{\sqrt{n}}$$

$$= 1.94 \cdot \frac{\sigma}{\sqrt{3}}.$$
Table 2.2. Values of $\alpha$ and $L$ for two-sided GMA charts with ARL 500 when $\mu = 0$ and values of ARL when $\mu \neq 0$

<table>
<thead>
<tr>
<th>$\sqrt{n \mu / \sigma}$</th>
<th>0.00</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>2.25</th>
<th>2.50</th>
<th>2.75</th>
<th>3.00</th>
<th>3.25</th>
<th>3.50</th>
<th>3.75</th>
<th>4.00</th>
<th>4.25</th>
<th>4.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>.90</td>
<td>.80</td>
<td>.70</td>
<td>.60</td>
<td>.50</td>
<td>.40</td>
<td>.30</td>
<td>.20</td>
<td>.10</td>
<td>0.0</td>
<td>0.25</td>
<td>0.50</td>
<td>0.75</td>
<td>1.00</td>
<td>1.25</td>
<td>1.50</td>
<td>1.75</td>
<td>2.00</td>
<td>2.25</td>
</tr>
<tr>
<td>$L$</td>
<td>2.85</td>
<td>2.97</td>
<td>3.03</td>
<td>3.06</td>
<td>3.07</td>
<td>3.08</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
<td>3.09</td>
</tr>
</tbody>
</table>
The procedure outlined above yields a GMA chart for controlling a process mean, which approximately satisfies the criteria proposed by Woodall. We earlier demonstrated that the GMA is an optimal estimator for the process mean under the steady model (2.3, 2.4). The ARLs used in the construction of a GMA chart presented here have been based on an i.i.d. model. Roberts (1959) compares several control chart procedures, including the X-bar and GMA charts under the i.i.d. model. He concludes that the standard Shewhart chart (the \( \alpha = 0 \) case of our GMA charts) is simpler than alternative charts and cannot be improved on in detecting relatively large changes from \( \mu = 0 \). In general, as interest is in early detection of smaller and smaller changes, the parameter \( \alpha \) appropriate for a GMA chart increases away from zero. The complexity of the GMA chart for the process mean relative to that of the standard Shewhart X-bar chart can be tolerated in cases where added sensitivity to relatively small changes in the process mean is desired. See Roberts (1966) for further comparisons involving the X-bar, GMA, Moving Average, and CUSUM charts.

E. A Generalized Kalman Filter for a Process Variance

Several approaches to monitoring a process variance have been suggested, including the Shewhart range (R), sample standard deviation (s), and sample variance (s^2) charts. Here, we will use a generalization of the steady model (2.1, 2.2) due to Smith (1979) to derive an alternative variance monitoring tool. An approach for obtaining (the i.i.d.)
ARLs for the proposed alternative will be developed, and comparisons to the Shewhart $s^2$ chart will be made.

Smith (1979) presents the steady model (2.1, 2.2) in terms of two stages:

**Stage 1:** $y_t = \theta_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2_\varepsilon)$

**Stage 2:** $\theta_t = \theta_{t-1} + \nu_t \quad \nu_t \sim N(0, \sigma^2_\nu)$

where the error components \{\varepsilon_t, \nu_t, t=1,2,\ldots\} are independent. Again, letting $y^t$ represent all the data through time $t$, the normal steady model gives

$$\theta_t | y^t \sim N(\hat{\theta}_t, q_t),$$

where

$$\hat{\theta}_t = (1-k_t)\hat{\theta}_{t-1} + k_t y_t$$

as in (2.5). Replacing $q_t$ with its limiting value we have approximately

$$\theta_t | y^t \sim N(\hat{\theta}_t, q)$$

and

$$\hat{\theta}_t = (1-k)\hat{\theta}_{t-1} + ky_t.$$

If we let $p(\theta_t | y^t)$ be the probability density function of $\theta_t$ given $y^t$, and $p(\theta_{t+1} | y^t)$ be the probability density function of $\theta_{t+1}$ given $y^t$,
then \( p(\theta_t | y^t) \) is the \( N(\hat{\theta}_t, q) \) density and from Stage 2 above, \( p(\theta_{t+1} | y^t) \) is the \( N(\hat{\theta}_t, q + \sigma^2_Y) \) density. In this case,

\[
p(\theta_{t+1} | y^t) \propto [p(\theta_t | y^t)]^k,
\]

where

\[
k = q/(q + \sigma^2_Y) < 1.
\]

Display (2.16) suggests that to generalize the steady model to distributions other than the normal, we can reformulate Stages 1 and 2 using (2.16). An equivalent way of writing Stage 1 in the normal model is

\[
y_t | \theta_t \sim N(\theta_t, \sigma^2) .
\]

This leaves the problem of how to update the distribution of \( \theta_{t-1} | y^{t-1} \) to that of \( \theta_t | y^{t-1} \). Once this update has been made, we can then determine

\[
p(\theta_t | y^t) \propto f(y_t | \theta_t) \cdot p(\theta_t | y^{t-1}) .
\]

For the normal steady model, the update is given by (2.16).

Using a Bayesian approach, Smith (1979) shows that for a particular loss structure, subject to certain reasonable criteria for choosing an appropriate model, an update of the form (2.16) is optimal. Smith
imposes the condition that the model must yield a decision which will remain constant over the time period \([t, t+1)\) when no more information is forthcoming. Also, the expected loss from taking this decision must always increase over this time interval. Note that in the normal steady model, the mean of \((\theta_{t+1}|y^t)\) is the same as that of \((\theta_t|y^t)\) while the variance is greater. Setting

\[ p(\theta_{t+1}|y^t) \propto [p(\theta_t|y^t)]^k \quad k < 1 \]

makes the density of \(\theta_{t+1}|y^t\) "flatter" than the density of \(\theta_t|y^t\), satisfying the second condition above. For many families of densities, which Smith calls linear expanding families, if density \(f(\theta)\) is in the chosen family, then for \(k < 1\), so is \(g(\theta)\alpha[f(\theta)]^k\). Examples of linear expanding families include the Normal, Gamma, Beta, Student t and Pareto families. A Gamma version of Smith's extension of the steady model can be used to develop an algorithm for recursively estimating a changing process variance based on successive \(s^2\) values.

Let \(s_t^2\) be the sample variance based on \(n\) observations from a \(N(0, \sigma_t^2)\) distribution at time \(t\),

\[ y_t = (n-1)s_t^2, \]

and

\[ \theta_t = 1/\sigma_t^2. \]
Following Smith, define a generalized steady model in stages as

**Stage 1:** \[ y_t | \theta_t \sim \text{Gamma} \left( \frac{n-3}{2}, \frac{\theta_t}{2} \right) \] (2.17)

and

**Stage 2:** \[ p(\theta_t | y^{t-1}) \propto [p(\theta_{t-1} | y^{t-1})]^k \text{ for } 0 < k < 1, \]

where \( p(\theta_t | y^S) \) is the probability density function of \( \theta_t \) given \( y^S \). We will take \( p(\theta_t | y^t) \) to be from the conjugate Gamma family for the sake of tractability and simplicity and will assume that the distribution of \( \theta_{t-1} | y^{t-1} \) has parameters \( \alpha_{t-1} \) and \( \beta_{t-1} \). That is, \( p(\theta_{t-1} | y^{t-1}) \) is the Gamma(\( \alpha_{t-1}, \beta_{t-1} \)) density. Then

\[ p(\theta_{t-1} | y^{t-1}) \propto \theta_{t-1}^{\alpha_{t-1}-1} e^{-\beta_{t-1} \theta}, \]

and from Stage 2:

\[ p(\theta_t | y^{t-1}) \propto \theta^{k\alpha_{t-1}-1} e^{-k\beta_{t-1} \theta}, \]

which is proportional to a Gamma(\( k\alpha_{t-1}, k\beta_{t-1} \)) density.

Note that the update (2.16) preserves the mode, \( \alpha_{t-1}/\beta_{t-1} \), while the variance increases. This is consistent with the conditions we wanted the update to satisfy. Now we have that

\[ p(\theta_t | y^t) \propto f(y_t | \theta_t) \cdot p(\theta_t | y^{t-1}) \]
\[
\alpha \left[ \frac{n-1}{\theta} \left( \frac{n-1}{2} \left( -1 - \frac{\theta}{2} y_t \right) \right) \right] \cdot \left[ \frac{k}{\theta} t - 1 \right] e^{-k \beta t - 1} \theta
\]

\[
\frac{n-1}{\theta} + \frac{k}{\theta} t - 1 - \frac{y_t}{2} + k \beta t - 1 \theta
\]

which is proportional to the Gamma\(\left( \frac{n-1}{2} + \frac{k}{\theta} t - 1, \frac{y_t}{2} + k \beta t - 1 \right)\) density.

So we have

\[
(\theta_t | y^t) \sim \text{Gamma}(\alpha_t, \beta_t), \tag{2.18}
\]

where

\[
\alpha_t = \frac{n-1}{2} + k \alpha_{t-1}
\]

and

\[
\beta_t = \frac{y_t}{2} + k \beta_{t-1}.
\]

The \(\alpha_t\) sequence does not depend on the data, and from Appendix B we have that

\[
\lim_{t \to \infty} \alpha_t = \frac{n-1}{2(1-k)}
\]

and the convergence is rapid. Also note that

\[
\beta_t = \frac{1}{2} \sum_{i=0}^{t} k^i y_{t-i} \text{ for } y_0 = \beta_0, \text{ some starting value.}
\]
Smith shows that under his proposed loss structure the optimal Bayes predictor of $\theta_t$ is the posterior mode of $\theta_t|y^t$. In the Gamma case, the posterior mode of $\theta_t|y^t$, $m_t$ satisfies

$$m_t^{-1} = \frac{\beta_t}{\alpha_t} = \frac{\frac{y_t}{2} + k\beta}{\frac{n-1}{2} + k\alpha_{t-1}}.$$

Using the limiting value of the sequence $\{\alpha_t\}$, for large $t$, $m_t^{-1}$ is approximately

$$\sum_{i=0}^{t} \frac{k^i y_{t-i}}{(\frac{n-1}{1-k})} = (1-k) \sum_{i=0}^{t} k^i \frac{i}{n-1}.$$

$$= (1-k) \sum_{i=0}^{t} k^i s^2_{t-1}.$$

Note that $m_t^{-1}$ is thus approximately a geometric moving average of the sample variances. We have that approximately

$$m_t^{-1} = km_{t-1}^{-1} + (1-k)s_t^2.$$  \hspace{1cm} (2.19)

Since $\theta_t = 1/\sigma_t^2$, instead of using $m_t$'s to monitor successive $1/\sigma_t^2$ values, we will use the $m_t^{-1}$'s given by (2.19) to monitor the process variances $\sigma_t^2$. 
It should be noted that throughout the above development of the recursive estimation procedure, the nature of the (marginal) joint distribution of the random variables \( \{ \theta_t \} \) has not been identified. Recall that in the normal steady model the parameter sequence \( \theta_t \) was a random walk sequence. For the Gamma version, we can express the joint distribution of the \( y_t \)'s and \( \theta_t \)'s as

\[
f(y_t, \theta_t) = f(y_t | \theta_t, \theta_{t-1}, y_{t-1}) \cdot f(\theta_t | \theta_{t-1}, y_{t-1})
\]

\[
\cdot f(y_{t-1} | \theta_{t-1}, \theta_{t-2}, y_{t-2}) \cdot f(\theta_{t-1} | \theta_{t-2}, y_{t-2})
\]

\[
\ldots f(y_2 | \theta_2, \theta_1, y_1) \cdot f(\theta_2 | \theta_1, y_1) \cdot f(y_1 | \theta_1) \cdot f(\theta_1)
\]

\[
\alpha \prod_{i=1}^{t} \left\{ \frac{\theta_i^{n-1}}{2} \frac{\theta_i^{n-3}}{2} e^{-\frac{\theta_i y_i}{2}} \right\} \cdot \left\{ \frac{\Gamma(k_{i-1}+1)}{\Gamma(k\alpha_{i-1}+1)} \frac{\theta_i^{k \alpha_{i-1}+1}}{2} e^{-\frac{\theta_i}{2} + k\beta_{i-1} \theta_i} \right\}
\]

\[
= \prod_{i=1}^{t} \frac{(k\beta_{i-1})}{\Gamma(k\alpha_{i-1}+1)} \frac{\theta_i^{n-1}}{2} \frac{\theta_i^{n-3}}{2} e^{-\frac{\theta_i y_i}{2}} e^{-\frac{\theta_i}{2} + k\beta_{i-1} \theta_i}
\]

\[
= \prod_{i=1}^{t} \frac{(k\beta_{i-1})}{\Gamma(k\alpha_{i-1}+1)} \frac{\theta_i^{n-1}}{2} \frac{\theta_i^{n-3}}{2} e^{-\frac{\theta_i}{2} + k\beta_{i-1} \theta_i}
\]

with \( \alpha_i, \beta_i \) defined by (2.18). So, in theory, we can write down an expression for the joint density of the \( \theta_t \)'s:
However, there appears to be no useful representation of this marginal joint distribution.

F. Average Run Lengths for a GMA Chart for Sample Variances

Consider using the estimation procedure for a process variance which follows from the Gamma extension of the steady model in a control charting context. The procedure is to plot successive $Q_t$ values, where $Q_t$ is defined by

$$Q_t = \alpha Q_{t-1} + (1-\alpha)s_t^2 \quad 0 < \alpha < 1, \ t = 1,2,\ldots$$

for $s_t^2$ the sample variance at time $t$ based on a sample of size $n$.

As in Section B, we have a GMA, but here a GMA of sample variances rather than sample means. The properties of a charting procedure based on such $Q_t$ do not appear to have been studied. Here we will obtain ARLs for this procedure assuming the $s_t^2$'s are i.i.d. Gamma $\left(\frac{n-3}{2}, \frac{n-1}{2\sigma^2}\right)$. The process will be deemed "out of control" (and we will conclude that the variance has increased) if $Q_t$ is too large. Without loss of generality, we will take the desired target variance to be $\sigma^2 = 1$. At time $t$, we will conclude the variance has increased if

$$Q_t > h, \text{ for } h \text{ some specified constant.}$$
As in Section C, let \( L(u) \) be the ARL given that the GMA starts at \( Q_0 = u \).

If \( x \sim \text{Gamma}\left(\frac{n-3}{2}, \frac{n-1}{2\sigma^2}\right) \),

\[
L(u) = 1 \cdot \Pr\left[\alpha u + (1-\alpha)x > h\right]
+ \int \left(1 + L(\alpha u + (1-\alpha)x)\right) f_\sigma(x)dx ,
\]

\( \{\alpha u + (1-\alpha)x < h\} \)

where \( f_\sigma(x) \) is the \( \text{Gamma}\left(\frac{n-3}{2}, \frac{n-1}{2\sigma^2}\right) \) density function. Thus,

\[
L(u) = 1 + \int L(\alpha u + (1-\alpha)x)f_\sigma(x)dx
\]

\( \{\alpha u + (1-\alpha)x < h\} \)

\[
= 1 + \int \frac{(h-\alpha u)(1-\alpha)}{\alpha u} f_\sigma(x)dx
= 1 + \int \frac{(1-\alpha)}{\alpha u} L(\alpha u + (1-\alpha)x)f_\sigma(x)dx
= 1 + \left(\frac{1}{1-\alpha}\right) \int_0^h L(t)f_\sigma\left(\frac{t-\alpha u}{1-\alpha}\right)dt .
\]

(2.20)

So, we have an integral equation of the form (2.9).

A quadrature rule as described in Section C was used to approximate \( L(u) \) in (2.20). Because of the form of \( f_\sigma(x) \), a trapezoidal rule was
used in the approximation. Values of $L(1)$, the ARL for the GMA starting at $Q_0 = 1$, were obtained for sample sizes $n = 2, 6, 10,$ and $20$. Note that a particular $L(1)$ value depends on the true process variance $\sigma^2$, the weighting constant $\alpha$, and the control limit $h$. Values of $L(1)$ for varying $(\alpha, h, \sigma^2)$ are given in Tables (2.3-2.6). The ARLs listed are consistent with simulation results which were obtained independently.

In Tables (2.3-2.6) the column headed $\sigma^2$ refers to the true variance. The values of $h$ listed correspond to different control limits, while the $\alpha$ values correspond to possible weighting constants used in the GMA. For example, with $n = 10$, from Table 2.5, a GMA with $\alpha = .50$ and $h = 1.88$ yields an in-control ARL of 221.

These ARL values can be used in the choice of appropriate $\alpha$ and $h$, the parameters of the GMA chart. And these values provide a way to compare a GMA chart for sample variances to the usual Shewhart $s^2$-chart. It is worth noting that, as with the GMA chart for the process mean, we could also have obtained expressions for expected values of any function of the run length, using the approach in Section C.

Duncan (1974) discusses the use of the most common approaches to monitoring a process variance, including the Shewhart $R$, $R$-bar, $s$, and $s^2$ charts. For small sample sizes he recommends the $R$-chart, while for large samples the $R$-bar, $s$, or $s^2$ charts are recommended. Grubbs (1947) has shown that the efficiency of the range is very good relative to the sample standard deviation for small sample sizes, say $n \leq 10$. As $n$ increases, the range, however, loses rapidly in efficiency.

Page (1963) and Hawkins (1981) have proposed different CUSUM schemes
Table 2.3. Average run length for geometric moving average of sample variances, n=2<sup>a</sup>

<table>
<thead>
<tr>
<th>σ&lt;sup&gt;2&lt;/sup&gt;</th>
<th>h=</th>
<th>.10</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
<th>.00</th>
<th>.10</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>5.34</td>
<td>5.70</td>
<td>7.03</td>
<td>9.41</td>
<td>18.10</td>
<td>6.25</td>
<td>7.93</td>
<td>8.61</td>
<td>14.37</td>
<td>46.05</td>
</tr>
<tr>
<td>1.25</td>
<td>3.78</td>
<td>3.94</td>
<td>4.34</td>
<td>6.53</td>
<td>12.87</td>
<td>3.94</td>
<td>5.06</td>
<td>6.59</td>
<td>9.61</td>
<td>23.62</td>
</tr>
<tr>
<td>1.50</td>
<td>3.12</td>
<td>4.03</td>
<td>3.66</td>
<td>4.86</td>
<td>8.46</td>
<td>4.89</td>
<td>5.13</td>
<td>5.18</td>
<td>6.79</td>
<td>13.05</td>
</tr>
<tr>
<td>1.75</td>
<td>2.61</td>
<td>2.98</td>
<td>3.39</td>
<td>3.33</td>
<td>6.31</td>
<td>3.20</td>
<td>3.12</td>
<td>4.50</td>
<td>6.34</td>
<td>9.76</td>
</tr>
<tr>
<td>2.00</td>
<td>2.58</td>
<td>3.08</td>
<td>3.04</td>
<td>3.58</td>
<td>4.86</td>
<td>3.34</td>
<td>3.18</td>
<td>3.54</td>
<td>4.39</td>
<td>7.66</td>
</tr>
<tr>
<td>3.00</td>
<td>2.03</td>
<td>2.23</td>
<td>2.14</td>
<td>2.72</td>
<td>3.56</td>
<td>2.75</td>
<td>2.63</td>
<td>2.84</td>
<td>2.91</td>
<td>5.12</td>
</tr>
<tr>
<td>4.00</td>
<td>1.95</td>
<td>2.12</td>
<td>1.96</td>
<td>2.22</td>
<td>2.51</td>
<td>2.17</td>
<td>2.35</td>
<td>2.43</td>
<td>2.52</td>
<td>3.52</td>
</tr>
<tr>
<td>5.00</td>
<td>1.81</td>
<td>1.79</td>
<td>1.78</td>
<td>2.30</td>
<td>2.34</td>
<td>1.88</td>
<td>1.70</td>
<td>2.02</td>
<td>2.23</td>
<td>2.93</td>
</tr>
</tbody>
</table>

<sup>a</sup>For Table 2.3, ARLs were approximated using 100 simulation runs. The estimated standard errors were roughly 10% of the estimated ARLs.
Table 2.3. (continued)

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h = 2.71$</th>
<th>$h = 3.84$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.0$</td>
<td>$0.10$</td>
<td>$0.25$</td>
</tr>
<tr>
<td>1.00</td>
<td>7.85</td>
<td>11.62</td>
</tr>
<tr>
<td>1.25</td>
<td>6.61</td>
<td>7.86</td>
</tr>
<tr>
<td>1.50</td>
<td>5.06</td>
<td>4.91</td>
</tr>
<tr>
<td>1.75</td>
<td>4.08</td>
<td>5.42</td>
</tr>
<tr>
<td>2.00</td>
<td>4.23</td>
<td>4.23</td>
</tr>
<tr>
<td>3.00</td>
<td>3.07</td>
<td>2.99</td>
</tr>
<tr>
<td>4.00</td>
<td>2.56</td>
<td>2.45</td>
</tr>
<tr>
<td>5.00</td>
<td>2.05</td>
<td>2.11</td>
</tr>
</tbody>
</table>
Table 2.3. (continued)

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h=0.0$</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.0</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>37.21</td>
<td>53.20</td>
<td>84.94</td>
<td>313.63</td>
<td>15017.1</td>
<td>103.71</td>
<td>146.92</td>
<td>305.96</td>
<td>1940.10</td>
<td>47548.9</td>
</tr>
<tr>
<td>1.25</td>
<td>24.97</td>
<td>29.89</td>
<td>36.79</td>
<td>135.85</td>
<td>1633.83</td>
<td>48.68</td>
<td>58.82</td>
<td>119.91</td>
<td>489.53</td>
<td>24631</td>
</tr>
<tr>
<td>1.50</td>
<td>14.37</td>
<td>16.67</td>
<td>21.40</td>
<td>56.05</td>
<td>459.81</td>
<td>31.83</td>
<td>33.00</td>
<td>63.67</td>
<td>184.25</td>
<td>5119.59</td>
</tr>
<tr>
<td>1.75</td>
<td>12.45</td>
<td>13.58</td>
<td>15.41</td>
<td>30.25</td>
<td>205.45</td>
<td>23.52</td>
<td>26.11</td>
<td>34.13</td>
<td>96.53</td>
<td>1227.66</td>
</tr>
<tr>
<td>2.00</td>
<td>8.50</td>
<td>10.99</td>
<td>13.62</td>
<td>24.93</td>
<td>108.17</td>
<td>13.82</td>
<td>16.74</td>
<td>24.59</td>
<td>52.95</td>
<td>426.89</td>
</tr>
<tr>
<td>3.00</td>
<td>5.60</td>
<td>5.31</td>
<td>5.62</td>
<td>9.60</td>
<td>22.24</td>
<td>7.17</td>
<td>7.15</td>
<td>8.76</td>
<td>22.00</td>
<td>59.39</td>
</tr>
<tr>
<td>4.00</td>
<td>4.17</td>
<td>4.44</td>
<td>4.60</td>
<td>8.25</td>
<td>12.58</td>
<td>5.09</td>
<td>5.31</td>
<td>5.80</td>
<td>11.28</td>
<td>26.14</td>
</tr>
<tr>
<td>5.00</td>
<td>3.65</td>
<td>3.18</td>
<td>3.03</td>
<td>4.09</td>
<td>8.89</td>
<td>4.09</td>
<td>3.87</td>
<td>4.84</td>
<td>6.54</td>
<td>13.97</td>
</tr>
</tbody>
</table>

Table 2.4. Average run length for geometric moving average of sample variances, n=6

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h=$</th>
<th>1.46</th>
<th>1.46</th>
<th>1.46</th>
<th>1.46</th>
<th>1.46</th>
<th>1.46</th>
<th>1.46</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.00</td>
<td>5.00</td>
<td>5.65</td>
<td>7.08</td>
<td>12.71</td>
<td>47.64</td>
<td>6.67</td>
<td>7.77</td>
</tr>
<tr>
<td></td>
<td>1.25</td>
<td>3.10</td>
<td>3.35</td>
<td>3.89</td>
<td>5.68</td>
<td>12.70</td>
<td>3.83</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>2.31</td>
<td>2.44</td>
<td>2.72</td>
<td>3.59</td>
<td>6.51</td>
<td>2.72</td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td>1.75</td>
<td>1.90</td>
<td>1.99</td>
<td>2.15</td>
<td>2.68</td>
<td>4.37</td>
<td>2.17</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>2.00</td>
<td>1.66</td>
<td>1.72</td>
<td>1.84</td>
<td>2.19</td>
<td>3.34</td>
<td>1.85</td>
<td>1.93</td>
</tr>
<tr>
<td></td>
<td>3.00</td>
<td>1.27</td>
<td>1.29</td>
<td>1.33</td>
<td>1.47</td>
<td>1.93</td>
<td>1.34</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>4.00</td>
<td>1.15</td>
<td>1.16</td>
<td>1.18</td>
<td>1.25</td>
<td>1.52</td>
<td>1.18</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
<td>1.09</td>
<td>1.10</td>
<td>1.11</td>
<td>1.16</td>
<td>1.33</td>
<td>1.11</td>
<td>1.12</td>
</tr>
</tbody>
</table>
Table 2.4. (continued)

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h = 1.85$</th>
<th>$h = 2.21$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.0$</td>
<td>$0.10$</td>
<td>$0.25$</td>
</tr>
<tr>
<td>1.00</td>
<td>10.01</td>
<td>12.18</td>
</tr>
<tr>
<td>1.25</td>
<td>5.18</td>
<td>5.91</td>
</tr>
<tr>
<td>1.50</td>
<td>3.44</td>
<td>3.79</td>
</tr>
<tr>
<td>2.00</td>
<td>2.15</td>
<td>2.29</td>
</tr>
<tr>
<td>3.00</td>
<td>1.45</td>
<td>1.50</td>
</tr>
<tr>
<td>4.00</td>
<td>1.24</td>
<td>1.27</td>
</tr>
<tr>
<td>5.00</td>
<td>1.15</td>
<td>1.17</td>
</tr>
</tbody>
</table>
Table 2.4. (continued)

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h = 2.57$</th>
<th>$h = 3.02$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.0$</td>
<td>.10</td>
<td>.25</td>
</tr>
<tr>
<td>1.00</td>
<td>40.23</td>
<td>57.11</td>
</tr>
<tr>
<td>1.25</td>
<td>14.74</td>
<td>18.75</td>
</tr>
<tr>
<td>1.50</td>
<td>7.81</td>
<td>9.33</td>
</tr>
<tr>
<td>1.75</td>
<td>5.08</td>
<td>5.84</td>
</tr>
<tr>
<td>2.00</td>
<td>3.74</td>
<td>4.19</td>
</tr>
<tr>
<td>3.00</td>
<td>1.96</td>
<td>2.09</td>
</tr>
<tr>
<td>4.00</td>
<td>1.50</td>
<td>1.56</td>
</tr>
<tr>
<td>5.00</td>
<td>1.31</td>
<td>1.35</td>
</tr>
</tbody>
</table>
Table 2.5. Average run length for geometric moving average of sample variances, n=10

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h_0$</th>
<th>1.36</th>
<th>1.48</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha=0.0$</td>
<td>.10</td>
<td>.50</td>
</tr>
<tr>
<td>1.00</td>
<td>5.00</td>
<td>5.71</td>
<td>7.31</td>
</tr>
<tr>
<td>1.25</td>
<td>2.72</td>
<td>2.94</td>
<td>3.39</td>
</tr>
<tr>
<td>1.50</td>
<td>1.93</td>
<td>2.03</td>
<td>2.23</td>
</tr>
<tr>
<td>1.75</td>
<td>1.57</td>
<td>1.62</td>
<td>1.74</td>
</tr>
<tr>
<td>2.00</td>
<td>1.37</td>
<td>1.41</td>
<td>1.48</td>
</tr>
<tr>
<td>3.00</td>
<td>1.10</td>
<td>1.11</td>
<td>1.14</td>
</tr>
<tr>
<td>4.00</td>
<td>1.04</td>
<td>1.04</td>
<td>1.05</td>
</tr>
<tr>
<td>5.00</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
</tr>
</tbody>
</table>
Table 2.5. (continued)

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h$=</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>3.00</th>
<th>4.00</th>
<th>5.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a=0.0$</td>
<td></td>
<td>10.01</td>
<td>4.39</td>
<td>2.72</td>
<td>2.02</td>
<td>1.66</td>
<td>1.19</td>
<td>1.07</td>
<td>1.03</td>
</tr>
<tr>
<td>$a=.10$</td>
<td></td>
<td>12.38</td>
<td>5.99</td>
<td>2.97</td>
<td>2.15</td>
<td>1.75</td>
<td>1.21</td>
<td>1.08</td>
<td>1.04</td>
</tr>
<tr>
<td>$a=.25$</td>
<td></td>
<td>18.44</td>
<td>6.33</td>
<td>3.49</td>
<td>2.43</td>
<td>1.92</td>
<td>1.26</td>
<td>1.11</td>
<td>1.05</td>
</tr>
<tr>
<td>$a=.50$</td>
<td></td>
<td>53.26</td>
<td>11.57</td>
<td>5.28</td>
<td>3.35</td>
<td>2.50</td>
<td>1.45</td>
<td>1.20</td>
<td>1.10</td>
</tr>
<tr>
<td>$a=.75$</td>
<td></td>
<td>790.42</td>
<td>41.36</td>
<td>11.96</td>
<td>6.42</td>
<td>4.41</td>
<td>2.15</td>
<td>1.58</td>
<td>1.33</td>
</tr>
<tr>
<td>$a=0.0$</td>
<td>$h=1.63$</td>
<td>20.05</td>
<td>7.16</td>
<td>3.89</td>
<td>2.65</td>
<td>2.05</td>
<td>1.29</td>
<td>1.12</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>$h=1.88$</td>
<td>26.94</td>
<td>8.58</td>
<td>4.41</td>
<td>2.90</td>
<td>2.20</td>
<td>1.33</td>
<td>1.14</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>$h=1.88$</td>
<td>47.19</td>
<td>12.01</td>
<td>5.54</td>
<td>3.54</td>
<td>2.53</td>
<td>1.43</td>
<td>1.18</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>$h=1.88$</td>
<td>221.15</td>
<td>28.41</td>
<td>9.75</td>
<td>5.29</td>
<td>3.59</td>
<td>2.89</td>
<td>1.98</td>
<td>1.59</td>
</tr>
</tbody>
</table>
Table 2.5 (continued)

<table>
<thead>
<tr>
<th>( \sigma^2 )</th>
<th>( h=2.1 )</th>
<th>( a=0.0 )</th>
<th>( .10 )</th>
<th>( .25 )</th>
<th>( .50 )</th>
<th>( .75 )</th>
<th>( 0.0 )</th>
<th>( .10 )</th>
<th>( .25 )</th>
<th>( .50 )</th>
<th>( .75 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>40.14</td>
<td>58.80</td>
<td>122.32</td>
<td>972.64</td>
<td>25000</td>
<td>99.67</td>
<td>164.06</td>
<td>431.63</td>
<td>7050.44</td>
<td>50000</td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>11.78</td>
<td>14.91</td>
<td>23.18</td>
<td>74.68</td>
<td>1550.35</td>
<td>22.80</td>
<td>31.19</td>
<td>56.47</td>
<td>290.70</td>
<td>20000</td>
<td></td>
</tr>
<tr>
<td>1.50</td>
<td>5.64</td>
<td>6.63</td>
<td>8.91</td>
<td>18.65</td>
<td>94.08</td>
<td>9.29</td>
<td>11.50</td>
<td>17.06</td>
<td>47.22</td>
<td>590.45</td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>3.52</td>
<td>3.96</td>
<td>4.94</td>
<td>8.43</td>
<td>23.68</td>
<td>5.18</td>
<td>6.06</td>
<td>8.07</td>
<td>16.27</td>
<td>71.86</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>2.56</td>
<td>2.81</td>
<td>3.35</td>
<td>5.15</td>
<td>11.45</td>
<td>3.48</td>
<td>3.93</td>
<td>4.92</td>
<td>8.44</td>
<td>23.45</td>
<td></td>
</tr>
<tr>
<td>3.00</td>
<td>1.42</td>
<td>1.49</td>
<td>1.63</td>
<td>2.12</td>
<td>3.74</td>
<td>1.63</td>
<td>1.73</td>
<td>1.96</td>
<td>2.71</td>
<td>5.09</td>
<td></td>
</tr>
<tr>
<td>4.00</td>
<td>1.17</td>
<td>1.20</td>
<td>1.27</td>
<td>1.51</td>
<td>2.41</td>
<td>1.25</td>
<td>1.30</td>
<td>1.41</td>
<td>1.78</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td>5.00</td>
<td>1.08</td>
<td>1.10</td>
<td>1.14</td>
<td>1.28</td>
<td>1.87</td>
<td>1.13</td>
<td>1.15</td>
<td>1.21</td>
<td>1.43</td>
<td>2.27</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.6. Average run length for geometric moving average of sample variances, n=20

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$h=1.26$</th>
<th>$h=1.33$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha=0.0$</td>
<td>.10</td>
</tr>
<tr>
<td>1.00</td>
<td>5.00</td>
<td>5.77</td>
</tr>
<tr>
<td>1.25</td>
<td>2.23</td>
<td>2.38</td>
</tr>
<tr>
<td>1.50</td>
<td>1.51</td>
<td>1.57</td>
</tr>
<tr>
<td>1.75</td>
<td>1.24</td>
<td>1.27</td>
</tr>
<tr>
<td>2.00</td>
<td>1.13</td>
<td>1.14</td>
</tr>
<tr>
<td>3.00</td>
<td>1.01</td>
<td>1.02</td>
</tr>
<tr>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 2.6. (continued)

<table>
<thead>
<tr>
<th>( \sigma^2 )</th>
<th>h=1.43</th>
<th>h=1.59</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a=0.0</td>
<td>.10</td>
</tr>
<tr>
<td>1.00</td>
<td>10.01</td>
<td>12.59</td>
</tr>
<tr>
<td>1.25</td>
<td>3.38</td>
<td>3.79</td>
</tr>
<tr>
<td>1.50</td>
<td>1.95</td>
<td>2.09</td>
</tr>
<tr>
<td>1.75</td>
<td>1.45</td>
<td>1.52</td>
</tr>
<tr>
<td>2.00</td>
<td>1.24</td>
<td>1.28</td>
</tr>
<tr>
<td>3.00</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td>4.00</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( \alpha = 0.0 )</td>
<td>0.10</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>------</td>
</tr>
<tr>
<td>1.00</td>
<td>40.20</td>
<td>60.80</td>
</tr>
<tr>
<td>1.25</td>
<td>8.19</td>
<td>10.14</td>
</tr>
<tr>
<td>1.50</td>
<td>3.46</td>
<td>3.95</td>
</tr>
<tr>
<td>1.75</td>
<td>2.12</td>
<td>2.33</td>
</tr>
<tr>
<td>2.00</td>
<td>1.59</td>
<td>1.70</td>
</tr>
<tr>
<td>3.00</td>
<td>1.08</td>
<td>1.10</td>
</tr>
<tr>
<td>4.00</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>5.00</td>
<td>1.00</td>
<td>1.01</td>
</tr>
</tbody>
</table>
for monitoring a process scale parameter. The method we propose is the GMA chart for sample variances.

Typically, a procedure for controlling the process variance will be used in combination with a control chart for the process mean. We will assume that considerations based on the properties of the control procedure for the mean have determined an appropriate sample size, n. Using our knowledge of the run length distribution of the GMA chart based on sample variances, we can design an appropriate control procedure for the process variance based on samples of that size n.

G. Choice of Parameters for a GMA Chart for a Process Variance

In designing a GMA chart to control process variance, we will again employ in-control, indifference, and out-of-control regions. The in-control region, $M_0$, consists of all values of $\sigma^2$ such that $0 < \sigma^2 < \sigma^2_0$, where $\sigma^2_0 > 1$. This corresponds to the set of shifts in the variance which will not require a process adjustment.

The region of indifference, I, contains all values of $\sigma^2$ such that $\sigma^2_0 < \sigma^2 < \sigma^2_1$. The out-of-control region, $M_1$, consists of all values of $\sigma^2$ such that $\sigma^2 > \sigma^2_1$, the region for which a control procedure should give an out-of-control signal quickly.

Again, it might be argued that if the process is operating well within specification limits, a slight change in the process variance should be tolerated. But clearly, any substantial percentage increase in the process variance should not be tolerated. In fact, for "tight
control," we can take \( \sigma_0^2 = \sigma_1^2 = 1 \), and look for a GMA procedure which is uniformly preferable over \( \sigma^2 > 1 \). Using Tables (2.3-2.6), two GMA procedures can be compared, based on their respective ARL profiles over the regions \( M_0 \) and \( M_1 \). Designing an appropriate GMA chart for the process variance consists of choosing chart parameters \((\alpha, h)\) to satisfy specific ARL conditions. The following sequence of steps will lead to a GMA charting scheme which approximately satisfies the proposed criteria.

**Step 1.** Choose the in-control, indifference, and out-of-control regions \( M_0, I, \) and \( M_1 \). Specify ARL values \( L_0, L_1 \) associated with \( \sigma^2 = \sigma_0^2 \) and \( \sigma^2 = \sigma_1^2 \), respectively.

**Step 2.** Construct a table of ARL values for the sample size in use. Tables (2.2-2.5) give such values for \( n=2, 6, 10, 20 \).

**Step 3.** From the table determined in Step 2, construct a table of values of \((\alpha, h)\) for which \( \text{ARL} = L_0 \) when \( \sigma^2 = \sigma_0^2 \).

**Step 4.** Looking through the table constructed in Step 3, find all \((\alpha, h)\) combinations for which \( \text{ARL} = L_1 \) when \( \sigma^2 = \sigma_1^2 \). If there is more than one such combination, take the one which appears to have the "best" ARL profile over \( M_0 \) and \( M_1 \), bearing in mind there may not exist a uniformly best choice.
Step 5. The control limit is h.

The procedure outlined above yields a GMA chart for controlling a process variance which approximately satisfies the conditions proposed by Woodall.

We earlier derived the GMA of sample variances as related to an approximately Bayes estimator in the steady model (2.17). The procedure for choosing parameters of a GMA chart presented here has been based on ARLs under an i.i.d. model. The properties of the GMA of sample variances under an i.i.d. model appear not to have been previously studied.

For small sample sizes, the efficiency of the range and the ease of its calculation have traditionally favored the R-chart for monitoring process variance. For large n, the s or $s^2$ chart has been favored due to the decline in efficiency of the range. Tables (2.3-2.6) can be used to compare the properties of the GMA chart of sample variances with the usual Shewhart $s^2$-chart. Note that the $\alpha = 0$ cases in the given tables correspond to the plotting of $s^2_t$ at time t. In general, as interest is in early detection of smaller and smaller percent changes in the process variance, the parameter $\alpha$ increases away from zero. While the GMA chart is more complicated than the Shewhart $s^2$-chart, it has greater sensitivity to small percent increase in the process variance. For detection of large relative changes, the Shewhart $s^2$-chart cannot be improved upon. These general properties of the GMA of sample variances are apparent from examination of tables (2.3-2.6), and the ARLs of the Shewhart $s^2$-chart.
III. ADAPTIVE KALMAN FILTERING AND STATISTICAL PROCESS CONTROL

In this chapter, we will extend the estimation and process control procedures for a mean, discussed in Chapter II, to the case of unknown variance terms which can change over time. The steady model (2.3, 2.4) involved constant variance terms that were assumed to be known and the estimation procedure discussed there converged to a GMA of sample means. Here we will estimate the variance terms from the data. In this case, the incorporation of the estimated variance components into the algorithm for estimating the process mean produces what is known as an adaptive Kalman filter.

For the case of constant but unknown variance components, several adaptive filtering algorithms have been proposed in the engineering literature. A review and bibliography is given by Mehra (1972). Sarris (1973) presents a maximum likelihood approach and Louv (1984) proposes a MINQUE algorithm for estimating variance components. The approach taken here will be to use maximum likelihood to estimate variance terms changing over time. For the resulting adaptive filter, two control charting techniques will be compared.

Also, the algorithm for monitoring a process variance, derived in Chapter II, will be extended to allow the posterior updating constant $k$ of (2.16) to vary over time and be unknown. A marginal likelihood function will then be used to develop an estimator for $k$ and then an adaptive filter for a process variance.
A. A Generalization of the Steady Model and Estimation of Variance Components

Consider the model

\[ y_{it} = \theta_t + \epsilon_{it} \quad i = 1,2,...,n \]  \hspace{1cm} (3.1)

\[ \theta_t = \theta_{t-1} + \eta_t \quad t = 1,2,... \]

where \( \theta_t \) is the true, unknown process mean at time \( t \), \( \{y_{it}, i=1,2,...,n\} \) are observed scalar outputs at time \( t \), each with mean \( \theta_t \), and \( \{\epsilon_{it}, i=1,2,...,n; t=1,2,...\} \) and \( \{\eta_t, t=1,2,...\} \) are independent sequences of independent normal random variables with zero mean and variances

\[ \text{Var}(\epsilon_{it}) = \sigma^2_{\epsilon t} \quad i = 1,2,...,n \]  \hspace{1cm} (3.2)

\[ \text{Var}(\eta_t) = \sigma^2_{\eta t} \quad t = 1,2,... \]

This model is a generalization of the steady model, with variance terms which can change over time. Again, this is a special case of the Kalman Filter model (1.1, 1.2) with

\[ F_t = 1_n, \text{ an } n\text{-vector of ones,} \]

\[ V_t = \sigma^2_{\epsilon t} I_n, \text{ } I_n \text{ nxn identity matrix,} \]

\[ G_t = 1, \]
and \( W_t = \sigma^2_{\nu t} \) \( t = 1, 2, \ldots \).

Letting \( \mathbf{y}_t \) represent the vector of outputs observed at time \( t \) and \( \mathbf{y}^t \) represent the set of all outputs observed through time \( t \), let \( \hat{\theta}_{t-1} \) and \( q_{t-1} \) be the mean and variance of \( \theta_{t-1} | \mathbf{y}^{t-1} \). Then,

\[ \theta_{t-1} | \mathbf{y}^{t-1} \sim N(\hat{\theta}_{t-1}, q_{t-1}) \]

and this distribution represents our knowledge about \( \theta_{t-1} \) at time \( t-1 \).

Using the same argument as in Chapter II, the posterior distribution for \( \theta_t \) after time \( t \) is

\[ \theta_t | \mathbf{y}^t \sim N(\hat{\theta}_t, q_t) \]

where

\[ \hat{\theta}_t = (1-k_t) \hat{\theta}_{t-1} + k_t \mathbf{y}_t \tag{3.3} \]

\[ k_t = \left( \frac{q_{t-1} + \sigma^2_{\nu t}}{q_{t-1} + \sigma^2_{\nu t} + \sigma^2_{\varepsilon t}/n} \right) \]

\[ q_t = (q_{t-1} + \sigma^2_{\nu t}) \left( \frac{\sigma^2_{\varepsilon t}/n}{\sigma^2_{\varepsilon t}/n + \sigma^2_{\nu t} + q_{t-1}} \right). \]

These are the Kalman filter recursive equations for estimating the \( \theta_t \) sequence, under the model described by (3.1, 3.2). If all the variance
terms are known, then $\hat{\theta}_t$ is again a weighted moving average, depending on the data only through $\bar{y}_1, ..., \bar{y}_t$. The Kalman weights $k_t$ can be calculated independent of the data, using (3.3).

We consider the case where the variance terms are not known and must be estimated from the data. Using (3.1, 3.2) we can express the model at time $t$ as

$$
\begin{pmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_t
\end{pmatrix} = \begin{pmatrix}
    1 & \theta_0 \\
    1 & \theta_0 \\
    \vdots \\
    1 & \theta_0
\end{pmatrix} \begin{pmatrix}
    1 & \nu_1 + \varepsilon_1 \\
    1 & \nu_1 + \nu_2 + \varepsilon_2 \\
    \vdots \\
    1 & \nu_1 + \nu_2 + \nu_3 + \varepsilon_t
\end{pmatrix},
$$

where $\theta_0$ is the starting level of the process mean, $\varepsilon_t$ is the vector of residuals at time $t$, and $1_n$ is an $n$-vector of ones.

So we have for the full data vector at time $t$,

$$
y^t \sim \text{MVN}_{nt}(\theta, \Sigma),
$$

where

$$
\begin{align*}
\theta &= 1_{nt} \theta_0 \\
\Sigma &= D \otimes 1_n + B \otimes 1_{n-1} \otimes 1_n^	op,
\end{align*}
$$

for $D = \text{diag}\{\sigma_i^2\}$.
and \[ B = \{ b_{ij} \} = \left\{ \frac{\min(i,j)}{\sum_{k=1}^{\min(i,j)} \sigma^2_{vk}} \right\}_{txt} \]

where \( \otimes \) denotes the usual matrix Kronecker product.

From the complicated form of \( V \) in (3.5), it is apparent that it is not feasible to use the full likelihood, at time \( t \), to estimate \( \sigma^2_{et} \) and \( \sigma^2_{vt} \). If we make the simplifying assumption that \( \sigma^2_{et} = \sigma^2_V \) and \( \sigma^2_{vt} = \sigma^2_V \), then we could use the maximum likelihood approach given by Sarris (1973) to estimate the unknown variances. A different possibility would be to retain the full generality of the models (3.4, 3.5), follow a suggestion in Phadke (1982), and consider using successive differences \( (\mathbf{y}_t - \mathbf{y}_{t-1}) \) in the estimation of \( \sigma^2_{et} \) and \( \sigma^2_{vt} \).

Using (3.1, 3.2) the distribution of \( z_t = \mathbf{y}_t - \mathbf{y}_{t-1} \) is

\[
z_t \sim \text{MVN}(0, \begin{pmatrix} \sigma^2_{et-1} + \sigma^2_{et} I_n & \sigma^2_{vt-1} I_{n-1} \\ \sigma^2_{vt-1} I_{n-1} & \sigma^2_{vt} I_{n-n} \end{pmatrix}).
\] (3.6)

It is the case that the likelihood function of \( z_t \) depends on the data through \( z_t'z_t \) and \( z_t'1 \), so that the maximum likelihood estimators of \( \sigma^2_{et} \) and \( \sigma^2_{vt} \) based on \( z_t \) will be functions of \( z_t'z_t \) and \( z_t'1 \). Since \( z_t \) is the vector of differences \( y_t - y_{t-1} \), its components will depend on the particular ordering of the data \( \{ y_{1t}, i=1,2,...,n \} \) in the vector \( y_t \).

And since the likelihood function depends on \( z_t'z_t \), maximum likelihood estimators of \( \sigma^2_{et} \) and \( \sigma^2_{vt} \) will depend upon the particular ordering of the data. This is not a desirable property, since from the form of
model (3.1) it is clear that the ordering of the data collected at time \( t \) should not affect the estimates of \( \sigma^2_{\varepsilon t} \) and \( \sigma^2_{\nu t} \).

To overcome the lack of symmetry that results from the use of (3.6) in estimating the variance components and obtain estimators of \( \sigma^2_{\varepsilon t} \) and \( \sigma^2_{\nu t} \) which are symmetric in the data, we will instead consider only the joint likelihood function of the following three independent statistics.

Let

\[
z_t = \bar{y}_t - \bar{y}_{t-1},
\]

the difference in successive sample means,

\[
d_t = \sum_{i=1}^{n} (y_{it} - \bar{y}_t)^2,
\]

and

\[
d_{t-1} = \sum_{i=1}^{n} (y_{it-1} - \bar{y}_{t-1})^2.
\]

Then, the joint likelihood function of \((z_t, d_{t-1}, d_t)\) for \( \sigma^2_{\nu t} > 0, \sigma^2_{\varepsilon t-1} > 0, \) and \( \sigma^2_{\varepsilon t} > 0 \) can be expressed

\[
L(z_t, d_{t-1}, d_t; \sigma^2_{\nu t}, \sigma^2_{\varepsilon t-1}, \sigma^2_{\varepsilon t}) = \frac{-z_t^2}{2\pi\left(\sigma^2_{\nu t} + \frac{\sigma^2_{\varepsilon t-1}}{n} + \frac{\sigma^2_{\varepsilon t}}{n}\right)^{-1/2}} \cdot 2(\sigma^2_{\nu t} + \frac{\sigma^2_{\varepsilon t-1}}{n} + \frac{\sigma^2_{\varepsilon t}}{n})^{-1/2} \cdot e^{-2(\sigma^2_{\nu t} + \frac{\sigma^2_{\varepsilon t-1}}{n} + \frac{\sigma^2_{\varepsilon t}}{n})}.\]
\[
\begin{align*}
\frac{n-1}{2} - 1 & \quad \frac{-d_t}{2\sigma^2_{\varepsilon t}} \\
2^{n-1} \cdot \frac{n-1}{2} & \quad e^{-d_t/\sigma^2_{\varepsilon t}} \\
(\sigma^2_{\varepsilon t})^{2} & \quad \frac{d_t}{2} \frac{n-1}{2} \\
(\sigma^2_{\varepsilon t})^{2} & \quad e^{-d_t/\sigma^2_{\varepsilon t}}. \\
\end{align*}
\]

Note that the statistics used to form (3.7) do not depend on the ordering of the data in \( \varepsilon_{t-1} \) and \( \varepsilon_t \).

The likelihood function above can be maximized over nonnegative choices of \( \sigma^2_{\varepsilon t} \) and positive choices of \( \sigma^2_{\varepsilon t-1} \) and \( \sigma^2_{\varepsilon t} \) to obtain estimates of \( \sigma^2_{\varepsilon t} \) and \( \sigma^2_{\varepsilon t-1} \), the variance components at time \( t \), for use in the recursions (3.3). Notice that

\[
\text{Log } L = c_0 - \frac{1}{2} \log(\sigma^2_{\varepsilon t} + \frac{\sigma^2_{\varepsilon t-1}}{n} + \frac{\sigma^2_{\varepsilon t}}{n}) - \frac{z_t^2}{2(\sigma^2_{\varepsilon t} + \frac{\sigma^2_{\varepsilon t-1}}{n} + \frac{\sigma^2_{\varepsilon t}}{n})}
- \frac{(n-1)\log(\sigma^2_{\varepsilon t-1})}{2} - \frac{d_t-1}{2\sigma^2_{\varepsilon t-1}}
- \frac{(n-1)\log(\sigma^2_{\varepsilon t})}{2} - \frac{d_t}{2\sigma^2_{\varepsilon t}}.
\]

Differentiating with respect to \( \sigma^2_{\varepsilon t} \),
So, \( \frac{\partial \log L}{\partial \sigma^2_{vt}} = 0 \) when \( \sigma^2_{vt} = z_t^2 - (\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t})/n \).

From the form of (3.8), for any fixed positive \( \sigma^2_{\varepsilon_{t-1}} \) and \( \sigma^2_{\varepsilon_t} \), \( L \) is maximized as a function of \( \sigma^2_{vt} > 0 \) by

\[
\hat{\sigma}^2_{vt} = \max[0, z_t^2 - (\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t})/n]
\]

To find overall maximizers of (3.7) it then suffices to separately maximize the formal expressions

\[
L' = \log L(z_t, d_{t-1}, d_t; 0, \sigma^2_{\varepsilon_{t-1}}, \sigma^2_{\varepsilon_t}) \tag{3.9}
\]

\[
= c_1 - \frac{1}{2} \log(\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t}) - \frac{nz^2_t}{2(\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t})}

- \frac{(n-1)}{2} \log(\sigma^2_{\varepsilon_{t-1}}) - \frac{d_{t-1}}{2\sigma^2_{\varepsilon_{t-1}}} - \frac{(n-1)}{2} \log(\sigma^2_{\varepsilon_t}) - \frac{d_t}{2\sigma^2_{\varepsilon_t}}
\]

and

\[
L'' = \log L(z_t, d_{t-1}, d_t; z_t^2 - (\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t})/n, \sigma^2_{\varepsilon_{t-1}}, \sigma^2_{\varepsilon_t}) \tag{3.10}
\]
over positive choices of \((\sigma_{\varepsilon_{t-1}}^2, \sigma_{\varepsilon_t}^2)\) and compare the resulting maxima.
(Of course for \((\sigma_{\varepsilon_{t-1}}^2 + \sigma_{\varepsilon_t}^2) > n \varepsilon_t^2\), \(L''\) is not a genuine likelihood.
But, should \(L''\) be maximized for \((\sigma_{\varepsilon_{t-1}}^2 + \sigma_{\varepsilon_t}^2) > n \varepsilon_t^2\), the fact that
\(L' > L''\) for such \((\sigma_{\varepsilon_{t-1}}^2, \sigma_{\varepsilon_t}^2)\) will guarantee that when compared, the
maximizers of \(L'\) and \(L''\) will lead to the maximizer of \(L\) in (3.7)).

To maximize \(L''\) in (3.10) notice that \(L'' \rightarrow -\infty\) as \(\sigma_{\varepsilon_{t-1}}^2 \rightarrow 0\) and/or
\(\sigma_{\varepsilon_t}^2 \rightarrow 0\), so that if \(L''\) has a maximum, it occurs in the interior of the
region defined by \(\sigma_{\varepsilon_{t-1}}^2 > 0\) and \(\sigma_{\varepsilon_t}^2 > 0\) and thus at a point where

\[
\frac{3L''}{3\sigma_{\varepsilon_{t-1}}^2} = \frac{3L''}{3\sigma_{\varepsilon_t}^2} = 0.
\]

Now, it is easy to show that

\[
\frac{3L''}{3\sigma_{\varepsilon_{t-1}}^2} = \frac{n-1}{2\sigma_{\varepsilon_{t-1}}^2} \left[ \frac{d_{t-1}/n-1}{\sigma_{\varepsilon_{t-1}}^2} - 1 \right]
\]

and

\[
\frac{3L''}{3\sigma_{\varepsilon_t}^2} = \frac{n-1}{2\sigma_{\varepsilon_t}^2} \left[ \frac{d_t/n-1}{\sigma_{\varepsilon_t}^2} - 1 \right],
\]

from which it is clear that \(L''\) is maximized by choice of \(\sigma_{\varepsilon_{t-1}}^2 = d_{t-1}/n-1\) and \(\sigma_{\varepsilon_t}^2 = d_t/n-1\).
To maximize $L'$ in (3.9), note that $L' \to \infty$ as $\sigma^2_{et-1} \to 0$ and/or $\sigma^2_{et} \to 0$, so that if $L'$ has a maximum, it occurs in the interior of the region defined by $\sigma^2_{et-1} > 0$ and $\sigma^2_{et} > 0$. Thus, the maximum occurs at a point where

$$\frac{\partial L'}{\partial \sigma^2_{et-1}} = \frac{\partial L'}{\partial \sigma^2_{et}} = 0.$$ 

Now,

$$\frac{\partial L'}{\partial \sigma^2_{et-1}} = \frac{-1}{2(\sigma^2_{et-1} + \sigma^2_{et})} + \frac{nz_t^2}{2(\sigma^2_{et-1} + \sigma^2_{et})^2} - \frac{n-1}{2\sigma^2_{et-1}} + \frac{d_{t-1}}{2\sigma^4_{et-1}}$$

and

$$\frac{\partial L'}{\partial \sigma^2_{et}} = \frac{-1}{2(\sigma^2_{et-1} + \sigma^2_{et})} + \frac{nz_t^2}{2(\sigma^2_{et-1} + \sigma^2_{et})^2} - \frac{n-1}{2\sigma^2_{et}} + \frac{d_{t}}{2\sigma^4_{et}},$$

from whence one has

$$2 \cdot \frac{\partial L'}{\partial \sigma^2_{et-1}} = \frac{nz_t^2}{2} - \frac{(\sigma^2_{et-1} + \sigma^2_{et})}{(\sigma^2_{et-1} + \sigma^2_{et})^2} + \frac{n-1}{\sigma^2_{et-1}} \left[ \frac{d_{t-1}/n-1}{\sigma^2_{et-1}} - 1 \right],$$

(3.12)

and

$$2 \cdot \frac{\partial L'}{\partial \sigma^2_{et}} = \frac{nz_t^2}{2} - \frac{(\sigma^2_{et-1} + \sigma^2_{et})}{(\sigma^2_{et-1} + \sigma^2_{et})^2} + \frac{n-1}{\sigma^2_{et}} \left[ \frac{d_{t}/n-1}{\sigma^2_{et}} - 1 \right].$$

Notice that from (3.12), if $\sigma^2_{et-1} + \sigma^2_{et} > nz_t^2$ and $L'$ has a maximum at $(\sigma^2_{et-1}, \sigma^2_{et})$, it follows that
\[
\frac{d_{t-1}}{n-1} > \sigma^2_{\varepsilon_{t-1}}
\]
and
\[
\frac{d_t}{n-1} > \sigma^2_{\varepsilon_t},
\]
and hence that \(d_{t-1}/n-1 + d_t/n-1 > n\zeta_t^2\). So, if \(d_{t-1}/n-1 + d_t/n-1 \leq n\zeta_t^2\), \(L'\) must be maximized for a \((\sigma^2_{\varepsilon_{t-1}}, \sigma^2_{\varepsilon_t})\) with \(\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t} \leq n\zeta_t^2\).

So consider two cases. First, if \(d_{t-1}/n-1 + d_t/n-1 \leq n\zeta_t^2\), then \(L'\) is maximized in the region where \(\sigma^2_{\varepsilon_{t-1}} + \sigma^2_{\varepsilon_t} \leq n\zeta_t^2\). For such \((\sigma^2_{\varepsilon_{t-1}}, \sigma^2_{\varepsilon_t})\), \(L' \leq L''\) so that the overall maximizers of \(L\) in (3.7) are those of \(L''\). That is, if \(d_{t-1}/n-1 + d_t/n-1 \leq n\zeta_t^2\), maximum likelihood estimates of \((\sigma^2_{\varepsilon_t}, \sigma^2_{\varepsilon_{t-1}}, \sigma^2_{\varepsilon_t})\) are

\[
\hat{\sigma}^2_{\varepsilon_{t-1}} = \frac{d_{t-1}}{n-1}, \quad (3.13)
\]
\[
\hat{\sigma}^2_{\varepsilon_t} = \frac{d_t}{n-1}, \quad (3.14)
\]
and
\[
\hat{\sigma}^2_{\nu_t} = z_t^2 - (\hat{\sigma}^2_{\varepsilon_{t-1}} + \hat{\sigma}^2_{\varepsilon_t})/n.
\]

In the case \(d_{t-1}/n-1 + d_t/n-1 > n\zeta_t^2\), \(L''\) (being maximized at \((\sigma^2_{\varepsilon_{t-1}}, \sigma^2_{\varepsilon_t}) = (d_{t-1}/n-1, d_t/n-1)\)) is maximized in the region where \(L'' < L'\) so that maximizers of \(L\) in (3.7) are those of \(L'\). Setting

\[
\frac{\partial L'}{\partial \sigma^2_{\varepsilon_{t-1}}} = 0 = \frac{\partial L'}{\partial \sigma^2_{\varepsilon_t}}
\]
leads to the expression
\[
\sigma^2_{\varepsilon t-1} = \frac{1}{2} \left[ \frac{\sigma^2_{\varepsilon t}}{d_{t/n-1}^2 - \sigma^2_{\varepsilon t}} \right] \left[ 1 + \frac{4(d_{t/n-1} - \sigma^2_{\varepsilon t})d_{t/n-1}}{\sigma^2_{\varepsilon t}} \right]^{1/2} - 1.
\]

(3.14)

Then, substituting this expression for \( \sigma^2_{\varepsilon t-1} \) into (3.9) leads to a function of \( \sigma^2_{\varepsilon t} \) that can be optimized using a one dimensional numerical search to produce the overall optimizers of \( L \) in (3.7). That is, if \( d_{t/n-1} + d_{t/n-1} > nz^2_t \), maximum likelihood estimates of \( (\sigma^2_{\nu t}, \sigma^2_{\varepsilon t-1}, \sigma^2_{\varepsilon t}) \) are

\[
\hat{\sigma}^2_{\nu t} = 0
\]

(3.15)

and \( \hat{\sigma}^2_{\varepsilon t-1} \) and \( \hat{\sigma}^2_{\varepsilon t} \) must be determined via a one dimensional numerical search to obtain \( \hat{\sigma}^2_{\varepsilon t} \) and substitution of the maximizing \( \hat{\sigma}^2_{\varepsilon t} \) into (3.14) to obtain \( \hat{\sigma}^2_{\varepsilon t-1} \).

**B. An Adaptive Kalman Filtering Algorithm for a Process Mean**

Using the maximum likelihood estimates described in (3.13, 3.15), the adaptive Kalman filter recursive equations become

\[
\hat{\theta}_t = (1-k_t)\hat{\theta}_{t-1} + k_t y_t
\]

(3.16)
\[ \hat{q}_t = (\hat{q}_{t-1} + \hat{\sigma}_v^2) \left( \frac{\hat{\sigma}_{\epsilon t}^2 / n}{\hat{\sigma}_{\epsilon t}^2 / n + \hat{\sigma}_v^2 + \hat{q}_{t-1}} \right). \]

Since the variance terms are estimated from the data, the posterior variance \( \hat{q}_t \) and the Kalman weight \( \hat{k}_t \) cannot be expected to converge to a limiting value. \( \theta_t \) is thus an adaptive exponentially weighted moving average.

Adaptive exponentially smoothed forecasting techniques have been extensively studied, with the primary concern being how to choose and continually adjust the weighting factor \( k^* \). One approach uses the behavior of past data in order to arrive at an optimum value, as in Rao and Shapiro (1970). Another approach, discussed by Trigg (1964), is to vary the weighting factor according to the value of a tracking signal which depends on previous forecasting errors.

In (3.3) the weighting factor \( k^* \) is determined from the Kalman filter recursions and knowledge of the variance terms. In (3.16), the variance terms which determine \( k^* \) are unknown, so maximum likelihood estimates are used to obtain an estimate of \( k^* \), \( \hat{k}_t \). In either case, there is no problem concerning how the weights \( k^* (\hat{k}_t) \) should be chosen, as the choice is a natural consequence of the Kalman structure.

Consider the weight \( \hat{k}_t \) in (3.16). During periods when the process mean is changing the most, a large weighting factor is desirable, to increase the influence of the most recent samples on \( \hat{\theta}_t \). If the estimate \( \hat{\sigma}_v^2 \) in (3.16) is large relative to \( \hat{q}_{t-1} \) and \( \hat{\sigma}_{\epsilon t}^2 \), then \( \hat{k}_t \) is large, and the most recent sample mean has the most influence. From
the form of the estimator $\hat{\sigma}_{\nu t}^2$ in (3.13, 3.15), we see that a large value of $(\bar{y}_t - \bar{y}_{t-1})^2$ will result in a large estimate of $\sigma_{\nu t}^2$ and hence a large weighting factor. So, if the process mean is changing significantly, it will result in large values for $(\bar{y}_t - \bar{y}_{t-1})^2$ and $\hat{\sigma}_{\nu t}^2$, as desired. During periods when the process is relatively stable, a small value for the weighting factor is desirable. In this case, small values of $(\bar{y}_t - \bar{y}_{t-1})^2$ will typically be observed, and the estimate of $\sigma_{\nu t}^2$ should be close to or equal zero. Then, the estimate $\hat{k}_t$ will also be small, and the current data will have less influence on $\hat{\theta}_t$, as desired.

Here we are concerned with the application of this adaptive estimation technique in control charting to monitor a process mean. One problem associated with an adaptive control chart is the choice of appropriate control limits or other signalling criterion. We will here consider both Shewhart-type fixed control limits and a type of adaptive signalling criterion, proposed by Hoadley (1981), which makes use of the successive estimated posterior variances $\hat{q}_t$.

C. Control Charting Alternatives for a Process Mean

From (3.16), after the data at time $t$ has been observed, we have estimates of the posterior mean and variance of $\hat{\theta}_t$, the current process mean. The posterior mean and variance of $\hat{\theta}_t$ given $y^t$ are estimated, respectively, by

$$\hat{\theta}_t = E(\theta_t | y^t) = (1-k_t)\hat{\theta}_{t-1} + k_t \bar{y}_t$$
and
\[ \hat{q}_t = \hat{V}(\theta_t | y^t) = (\hat{q}_{t-1} + \hat{q}^2_{vt}) \left( \frac{\hat{\sigma}^2_n}{\hat{\sigma}^2_n + \hat{\sigma}^2_{vt} + q_{t-1}} \right). \]

Using an idea suggested by Hoadley (1981), a box and whisker plot can be made each period, in addition to calculation of the adaptive point estimate \( \hat{\theta}_t \). The box plot will be used to make a graphical representation of the estimated posterior distribution of the current process mean \( \theta_t \). An '0' at the middle of the box will represent the estimated posterior mean of the process level. A 'Y' on the chart at time \( t \) will represent the observed sample mean at time \( t \). The top and bottom of the box and the end of the whiskers will be drawn a fixed number of estimated posterior standard deviations of \( \theta_t \) from \( \hat{\theta}_t \). When either the top or bottom whisker fails to cross the target value, the process is deemed "out of control." A warning signal is given if the box fails to cross the target line. An illustration of the charting procedure is given in Figure 3.1.

In Figure 3.1, the box and whiskers could extend, respectively, 2 and 3 posterior standard deviations from \( \hat{\theta}_t \). Box plot 1 represents a process in-control, box plot 2 signals a warning that the process mean may be wandering off target, while box plot 3 signals that the process is out-of-control. Figure 3.2 illustrates the adaptive algorithm at work in box plot form, using simulated \((\bar{y}_t, s^2_t)\) pairs, for \( t = 1, 2, \ldots, 10 \).

Alternative to an adaptive (box plot type) signalling scheme is the use of Shewhart-type control limits consisting of fixed boundaries.
$\theta_0$ - Target value

$\hat{\theta}_t$ - Estimate of process mean, $\hat{\theta}_t$

$\bar{y}_t$ - Sample mean, $\bar{y}_t$

Figure 3.1. Box and whisker plots for a process mean
$O$ - Estimate of posterior mean
$Y$ - Sample mean
Box and whiskers extend 2 and 3 estimated posterior standard deviations respectively, from $O$.

Figure 3.2 Adaptive algorithm for process mean
When the estimate $\hat{\theta}_t$ falls outside these boundaries, the process will be deemed out of control.

The ARL properties of both of these control schemes based on the adaptive Kalman Filter were studied, again using an i.i.d. model for the variables \(\{y_{it}, i=1,2,\ldots,n; t=1,2,\ldots\}\). Because of the adaptive nature of the procedures, analytical expressions for the ARLs could not be derived. Simulations were performed to obtain empirical estimates of the ARLs. Tables (3.1-3.4) and (3.5-3.8) contain these ARLs for the two control schemes, for samples of size \(n=2, 6, 10,\) and \(20\). These ARLs were obtained via simulation of the i.i.d. model. The standard errors of the simulation estimates were roughly 10% of the estimated means. In Tables (3.1-3.4), the row heading \(L\) refers to the fixed control limit given in multiples of the standard deviation of the sample mean. In Tables (3.5-3.8), \(L\) refers to the box plot type control limit in terms of multiples of \(\sqrt{\hat{q}_t}\). In Tables (3.1-3.8), column heading \(\mu\) refers to the true fixed mean from the i.i.d. model, measured in units of the standard deviation of the \(\bar{y}_t\)'s. The algorithms were started with $\hat{\theta}_0 = 0$ and $\hat{q}_0 = \sigma^2$, the variance of the \(y_{it}\)'s from the i.i.d. model.

For sample size \(n=2\), Table 3.5 shows that the "in-control" ARLs associated with the box-whisker approach are much too small to be practical, even using whiskers as long as 8 posterior standard deviations. This problem is the same as would be experienced with Shewhart plotting of successive t statistics with such small \(n\). Our box-whisker approach corresponds to the plotting of "Bayesian" t statistics since
Table 3.1. ARL values: Adaptive filter for process mean, Shewhart control limits, n=2

<table>
<thead>
<tr>
<th>μ</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>20.30</td>
<td>38.40</td>
<td>174.80</td>
<td>607.70</td>
<td>3755.00</td>
</tr>
<tr>
<td>.25</td>
<td>14.30</td>
<td>30.60</td>
<td>84.10</td>
<td>363.90</td>
<td>2219.00</td>
</tr>
<tr>
<td>.50</td>
<td>5.00</td>
<td>22.00</td>
<td>53.50</td>
<td>166.80</td>
<td>716.20</td>
</tr>
<tr>
<td>.75</td>
<td>2.60</td>
<td>11.00</td>
<td>24.60</td>
<td>67.50</td>
<td>202.90</td>
</tr>
<tr>
<td>1.00</td>
<td>1.90</td>
<td>6.40</td>
<td>15.70</td>
<td>31.70</td>
<td>102.50</td>
</tr>
<tr>
<td>1.50</td>
<td>1.10</td>
<td>2.74</td>
<td>5.44</td>
<td>9.78</td>
<td>22.36</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>1.69</td>
<td>2.83</td>
<td>3.61</td>
<td>7.38</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>1.15</td>
<td>1.47</td>
<td>1.77</td>
<td>3.20</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.03</td>
<td>1.22</td>
<td>1.28</td>
<td>1.54</td>
</tr>
</tbody>
</table>

L = standard deviations of sample mean

Table 3.2. ARL values: Adaptive filter for process mean, Shewhart control limits, n=6

<table>
<thead>
<tr>
<th>μ</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>20.30</td>
<td>57.90</td>
<td>201.60</td>
<td>969.70</td>
<td>4765.00</td>
</tr>
<tr>
<td>.25</td>
<td>12.90</td>
<td>28.60</td>
<td>93.30</td>
<td>256.00</td>
<td>1268.00</td>
</tr>
<tr>
<td>.50</td>
<td>5.70</td>
<td>13.60</td>
<td>20.30</td>
<td>77.00</td>
<td>237.30</td>
</tr>
<tr>
<td>.75</td>
<td>3.10</td>
<td>5.20</td>
<td>10.50</td>
<td>23.80</td>
<td>52.90</td>
</tr>
<tr>
<td>1.00</td>
<td>1.50</td>
<td>2.30</td>
<td>4.20</td>
<td>7.20</td>
<td>17.30</td>
</tr>
<tr>
<td>1.50</td>
<td>1.10</td>
<td>1.18</td>
<td>1.25</td>
<td>1.64</td>
<td>3.03</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.03</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.3. ARL values: Adaptive filter for process mean, Shewhart control limits, n=10

<table>
<thead>
<tr>
<th>μ</th>
<th>L = standard deviations of sample mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>.00</td>
<td>21.40</td>
</tr>
<tr>
<td>.25</td>
<td>9.90</td>
</tr>
<tr>
<td>.50</td>
<td>3.70</td>
</tr>
<tr>
<td>.75</td>
<td>1.80</td>
</tr>
<tr>
<td>1.00</td>
<td>1.20</td>
</tr>
<tr>
<td>1.50</td>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 3.4. ARL values: Adaptive filter for process mean, Shewhart control limits, n=20

<table>
<thead>
<tr>
<th>μ</th>
<th>L = standard deviations of sample mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>.00</td>
<td>22.40</td>
</tr>
<tr>
<td>.25</td>
<td>7.10</td>
</tr>
<tr>
<td>.50</td>
<td>2.00</td>
</tr>
<tr>
<td>.75</td>
<td>1.10</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1.50</td>
<td>1.00</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.5. ARL values: Adaptive filter for process mean, box plot control limits, n=2

<table>
<thead>
<tr>
<th>μ</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>3.97</td>
<td>4.99</td>
<td>6.34</td>
<td>7.87</td>
<td>8.41</td>
<td>11.11</td>
<td>13.41</td>
</tr>
<tr>
<td>.25</td>
<td>3.45</td>
<td>4.97</td>
<td>6.27</td>
<td>8.29</td>
<td>9.60</td>
<td>10.36</td>
<td>13.60</td>
</tr>
<tr>
<td>.50</td>
<td>3.11</td>
<td>4.04</td>
<td>5.76</td>
<td>6.09</td>
<td>7.44</td>
<td>8.42</td>
<td>10.53</td>
</tr>
<tr>
<td>.75</td>
<td>2.81</td>
<td>3.60</td>
<td>4.00</td>
<td>4.88</td>
<td>6.21</td>
<td>6.65</td>
<td>7.60</td>
</tr>
<tr>
<td>1.00</td>
<td>2.13</td>
<td>2.85</td>
<td>3.25</td>
<td>4.35</td>
<td>4.65</td>
<td>5.59</td>
<td>5.93</td>
</tr>
<tr>
<td>1.50</td>
<td>1.58</td>
<td>1.94</td>
<td>2.43</td>
<td>2.58</td>
<td>3.19</td>
<td>3.60</td>
<td>3.94</td>
</tr>
<tr>
<td>2.00</td>
<td>1.31</td>
<td>1.47</td>
<td>1.83</td>
<td>2.25</td>
<td>2.33</td>
<td>2.85</td>
<td>3.13</td>
</tr>
<tr>
<td>2.50</td>
<td>1.12</td>
<td>1.39</td>
<td>1.61</td>
<td>1.77</td>
<td>2.14</td>
<td>2.47</td>
<td>2.71</td>
</tr>
<tr>
<td>3.00</td>
<td>1.04</td>
<td>1.29</td>
<td>1.38</td>
<td>1.46</td>
<td>2.01</td>
<td>1.95</td>
<td>2.33</td>
</tr>
</tbody>
</table>

Table 3.6. ARL values: adaptive filter for process mean, box plot control limits, n=6

<table>
<thead>
<tr>
<th>μ</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>12.62</td>
<td>31.99</td>
<td>103.79</td>
<td>234.79</td>
<td>596.54</td>
<td>1222.00</td>
<td>2001.00</td>
</tr>
<tr>
<td>.25</td>
<td>6.62</td>
<td>13.92</td>
<td>36.78</td>
<td>96.26</td>
<td>218.18</td>
<td>415.30</td>
<td>1026.00</td>
</tr>
<tr>
<td>.50</td>
<td>3.90</td>
<td>5.67</td>
<td>10.11</td>
<td>20.10</td>
<td>40.04</td>
<td>81.77</td>
<td>183.70</td>
</tr>
<tr>
<td>.75</td>
<td>2.25</td>
<td>3.25</td>
<td>5.43</td>
<td>8.75</td>
<td>13.79</td>
<td>24.63</td>
<td>42.32</td>
</tr>
<tr>
<td>1.00</td>
<td>1.48</td>
<td>2.27</td>
<td>3.14</td>
<td>4.31</td>
<td>6.92</td>
<td>9.61</td>
<td>16.36</td>
</tr>
<tr>
<td>1.50</td>
<td>1.17</td>
<td>1.42</td>
<td>1.85</td>
<td>2.22</td>
<td>2.93</td>
<td>3.92</td>
<td>4.74</td>
</tr>
<tr>
<td>2.00</td>
<td>1.01</td>
<td>1.11</td>
<td>1.41</td>
<td>1.54</td>
<td>1.97</td>
<td>2.37</td>
<td>2.97</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>1.03</td>
<td>1.10</td>
<td>1.24</td>
<td>1.57</td>
<td>1.94</td>
<td>2.03</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.09</td>
<td>1.25</td>
<td>1.41</td>
<td>1.51</td>
</tr>
</tbody>
</table>
### Table 3.7. ARL values: Adaptive filter for process mean, box plot control limits, n=10

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>18.04</td>
<td>77.49</td>
<td>308.60</td>
<td>1566.00</td>
<td>4354.00</td>
</tr>
<tr>
<td>.25</td>
<td>5.70</td>
<td>15.49</td>
<td>47.60</td>
<td>136.20</td>
<td>715.70</td>
</tr>
<tr>
<td>.50</td>
<td>2.53</td>
<td>4.41</td>
<td>7.73</td>
<td>18.00</td>
<td>33.90</td>
</tr>
<tr>
<td>.75</td>
<td>1.70</td>
<td>2.40</td>
<td>3.59</td>
<td>5.51</td>
<td>9.68</td>
</tr>
<tr>
<td>1.00</td>
<td>1.17</td>
<td>1.59</td>
<td>2.29</td>
<td>3.24</td>
<td>4.69</td>
</tr>
<tr>
<td>1.50</td>
<td>1.03</td>
<td>1.09</td>
<td>1.28</td>
<td>1.57</td>
<td>2.13</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.03</td>
<td>1.15</td>
<td>1.35</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.14</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Table 3.8. ARL values: Adaptive filter for process mean, box plot control limits, n=20

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>17.99</td>
<td>117.34</td>
<td>1141.00</td>
<td>6744.00</td>
<td>40519.00</td>
</tr>
<tr>
<td>.25</td>
<td>3.94</td>
<td>8.77</td>
<td>28.67</td>
<td>82.81</td>
<td>458.30</td>
</tr>
<tr>
<td>.50</td>
<td>1.64</td>
<td>2.64</td>
<td>4.42</td>
<td>7.47</td>
<td>13.10</td>
</tr>
<tr>
<td>.75</td>
<td>1.20</td>
<td>1.48</td>
<td>2.07</td>
<td>3.05</td>
<td>4.35</td>
</tr>
<tr>
<td>1.00</td>
<td>1.03</td>
<td>1.14</td>
<td>1.35</td>
<td>1.73</td>
<td>2.43</td>
</tr>
<tr>
<td>1.50</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
<td>1.07</td>
<td>1.36</td>
</tr>
<tr>
<td>2.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
we are essentially plotting values of $\hat{\theta}_t / \sqrt{\hat{q}_t}$. In the case $n=2$, the
Shewhart-type constant control limits seem preferable to the box-whisker
adaptive limits.

For sample sizes $n=6, 10, 20$, the two approaches can be compared by
setting "in-control" ARLs equal, then comparing "out-of-control" ARLs
over the various shifts in the mean. In each case, the box-whisker
chart appears to respond more quickly to shifts in the mean less than 1
standard deviation, suggesting a greater sensitivity to small shifts.
The larger shifts appear to be detected more quickly by the fixed limit
control chart. This can be seen from examination of Tables (3.1-3.8).

Evaluating the box-whisker approach under an i.i.d. model, the
successive estimated posterior variances tend to decrease until an out-
of-control signal is given. For small shifts in the process mean, this
apparently happens faster than the successive estimates $\hat{\theta}_t$ can move
outside a corresponding fixed control limit. For large shifts in the
mean, the estimate $\hat{\theta}_t$ jumps quickly outside the control limit, before
the posterior variance has become small enough to cause an out-of-control
signal. So, depending on the type of sensitivity desired, for $n=6, 10,$
20, either of the approaches might be preferred.

Both of the approaches described above can be compared to a Shewhart
t-statistic control chart, in which successive values of

$$Q_t = \sqrt{n} \bar{y}_t / s_t$$
are plotted, where $\bar{y}_t$ and $s_t$ are the sample mean and sample standard deviation, respectively, observed at time $t$. Under an i.i.d. model, $\bar{y}_t \sim \mathcal{N}(\mu, \sigma^2/n)$, $(n-1)s^2_t/\sigma^2 \sim \chi^2_{n-1}$, and $Q_t$ has a noncentral $t$ distribution with noncentrality parameter $\sqrt{n\mu}/\sigma$. Both adaptive control procedures appear to be superior to the $t$ statistic procedure in terms of ARLs corresponding to small values of the process mean. For large values of the process mean, the ARLs corresponding to Shewhart plotting of the $t$-statistic are similar to those corresponding to Shewhart plotting of the adaptive filter estimate. Table 3.9 gives the ARL values of the $t$ statistic procedure, for sample size $n=6$ and can be compared to Tables 3.2 and 3.6.

To construct adaptive control charts for monitoring a process mean, the procedure outlined in Chapter II is recommended.

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
<th>6.0</th>
<th>7.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>9.81</td>
<td>33.22</td>
<td>96.85</td>
<td>243.57</td>
<td>541.41</td>
<td>1089.43</td>
</tr>
<tr>
<td>.25</td>
<td>6.66</td>
<td>19.98</td>
<td>54.41</td>
<td>131.50</td>
<td>285.24</td>
<td>564.97</td>
</tr>
<tr>
<td>.50</td>
<td>3.49</td>
<td>8.73</td>
<td>21.47</td>
<td>48.92</td>
<td>102.34</td>
<td>198.06</td>
</tr>
<tr>
<td>.75</td>
<td>2.08</td>
<td>4.35</td>
<td>9.60</td>
<td>20.49</td>
<td>41.16</td>
<td>77.55</td>
</tr>
<tr>
<td>1.00</td>
<td>1.47</td>
<td>2.57</td>
<td>5.04</td>
<td>9.98</td>
<td>19.12</td>
<td>34.90</td>
</tr>
<tr>
<td>1.50</td>
<td>1.07</td>
<td>1.37</td>
<td>2.09</td>
<td>3.49</td>
<td>5.97</td>
<td>10.08</td>
</tr>
<tr>
<td>2.00</td>
<td>1.01</td>
<td>1.07</td>
<td>1.32</td>
<td>1.84</td>
<td>2.75</td>
<td>4.22</td>
</tr>
<tr>
<td>2.50</td>
<td>1.00</td>
<td>1.01</td>
<td>1.08</td>
<td>1.29</td>
<td>1.69</td>
<td>2.34</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.09</td>
<td>1.27</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Recall the generalization of the steady model for successive sample variances (2.16). Stage 2 of the model consisted of updating $p(\theta_{t-1} | y^{t-1})$ to $p(\theta_t | y^{t-1})$ by

$$p(\theta_t | y^{t-1}) \propto [p(\theta_{t-1} | y^{t-1})]^k \quad 0 < k < 1$$

producing a "flattened" version of $p(\theta_{t-1} | y^{t-1})$. The value of $k$ was assumed to be constant over time, and ended up appearing as a weight in a GMA of sample variances. Now, we consider letting $k$ vary with $t$, estimating these $k_t$'s and ultimately using these data based weights in an adaptive filtering algorithm for sample variances. (This corresponds to what we did earlier in the adaptive Kalman filter (3.16) for the process mean.)

So, stage 2 of the gamma generalized steady model becomes

$$p(\theta_t | y^{t-1}) \propto [p(\theta_{t-1} | y^{t-1})]^k_t \quad k_t \in (0,1) \quad (3.17)$$

and we are left with the problem of estimating $k_t$. At time $t$, we have for $y_t = (n-1)s_t^2$, $v = \frac{n-3}{2}$, and $\theta_t = \frac{1}{\sigma_t^2}$,

$$y_t | \theta_t \sim \text{Gamma}(v, \frac{\theta_t}{2})$$
and $\theta_t \mid (y_{t-1}, k_t) \sim \text{Gamma}(k_t \alpha_{t-1}, k_t \beta_{t-1})$, where $s_t^2$ is the sample variance at time $t$. Then, given $y_{t-1}$,

$$f(y_t \mid k_t) = \int_0^\infty f(y_t \mid \theta)p(\theta \mid y_{t-1}, k_t)d\theta$$

$$= \int_0^\infty \frac{\theta^{v+1}y_t^{v}e^{-\theta y_t/2}}{\Gamma(v+1)} \frac{(k_t \beta_{t-1})^{k_t \alpha_{t-1}+1}}{\Gamma(k_t \alpha_{t-1}+1)} \theta^{k_t \alpha_{t-1}+1}e^{-k_t \beta_{t-1} \theta}d\theta$$

$$= \frac{y_t^{v(k_t \beta_{t-1})^{k_t \alpha_{t-1}+1}}}{\Gamma(v+1)\Gamma(k_t \alpha_{t-1}+1)} \int_0^\infty \theta^{v+k_t \alpha_{t-1}+1}e^{-\theta (y_t/2 + k_t \beta_{t-1})}d\theta$$

$$= \frac{y_t^{v(k_t \beta_{t-1})^{k_t \alpha_{t-1}+1}}}{\Gamma(v+1)\Gamma(k_t \alpha_{t-1}+1)} \cdot \frac{\Gamma(v+k_t \alpha_{t-1}+2)}{(y_t/2 + k_t \beta_{t-1})^{v+k_t \alpha_{t-1}+2}}. \quad (3.18)$$

Since $\alpha_{t-1}$ and $\beta_{t-1}$, or estimates of them are available when $y_t$ is observed, (3.18) can be used as a marginal likelihood function for $y_t$, which depends only on the unknown $k_t$. So, $f(y_t \mid k_t)$ can be maximized over $(0,1)$ as a function of $k_t$. Once an estimate for $k_t$ has been determined, the update in (3.17) can be made, and as in Chapter II, we can derive the estimated posterior distribution

$$\theta_t \mid y_t \sim \text{Gamma}(\hat{\alpha_t}, \hat{\beta_t}),$$
where
\[ \hat{\alpha}_t = \frac{n-1}{2} + k_t \hat{\alpha}_{t-1} \]
and
\[ \hat{\beta}_t = \frac{y_t}{2} + k_t \hat{\beta}_{t-1} . \]

Again, in the Gamma case, the estimated posterior mode of \( \theta_t | y^t, \hat{m}_t \) satisfies

\[ \hat{\alpha}_t = \frac{n-1}{2} + k_t \hat{\alpha}_{t-1} \]
\[ \hat{\beta}_t = \frac{y_t}{2} + k_t \hat{\beta}_{t-1} . \]

And since \( \hat{\theta}_t = \frac{1}{\hat{\sigma}_t^2} \), instead of using \( \hat{m}_t \)'s to monitor successive \( 1/\sigma_t^2 \) values, we will use the adaptive \( \hat{m}_t \)'s to monitor the process variances \( \sigma_t^2 \).

In equations (3.19), note that a value of \( \hat{k}_t \) close to zero corresponds to \( \hat{\rho}_t \) near one. In this case, the update (3.17) \( p(\theta_t | y^{t-1}) \) is an extremely flat version of \( p(\theta_{t-1} | y^{t-1}) \), and the estimator \( \hat{m}_t \).
places most of its weight on the current data, \( s_t^2 \). If \( \hat{k}_t \) is close to one, then \( \hat{\rho}_t \) is less than one, and can be close to zero. Then, the prior \( p(\theta_t | y^{t-1}) \) is essentially a copy of \( p(\theta_{t-1} | y^{t-1}) \), and the estimator \( \hat{m}_t^{-1} \) places weight on both the current and past data.

To determine \( \hat{k}_t \), the likelihood (3.18) must be maximized numerically at each stage. Plots of \( f(y_t | k_t) \) were made for various values of \( y_t \), \( \alpha_{t-1} \), and \( \beta_{t-1} \). Figures 3.3 and 3.4 illustrate the two forms of the likelihood \( L(*) \) as a function of \( k_t \) on the interval \( (0,1) \) that were obtained in this graphical investigation.

![Figure 3.3. L(k), \( \hat{k} < 1 \)](image)

![Figure 3.4. L(k), \( \hat{k} = 1 \)](image)
A check of the derivative of $L(k)$ at $k=1$ determines whether the maximum occurs inside or outside the interval $(0,1)$. If Figure 3.4 applies, the point of maximum is taken to be $k=1$. Otherwise, a numerical search is made to find the point of maximum inside the interval.

The complete adaptive filtering algorithm for estimating the process variance $\sigma_t^2$, is

$$m_t^{-1} = \rho_t s_t^2 + (1-\rho_t)m_{t-1}^{-1} \tag{3.20}$$

where

$$\hat{\rho}_t = (n-1)/(n-1+2k_{t-1})$$

$$\hat{\beta}_t = \frac{(n-1)s_t^2}{2} + k_t \hat{\beta}_{t-1}$$

$$\hat{\alpha}_t = \frac{n-1}{2} + k_t \hat{\alpha}_{t-1}$$

and $k_t$ is found by maximizing (3.18) over $k_t \in (0,1]$.

At time $t$, once $s_t^2$ has been observed, the estimated posterior distribution for $\frac{1}{\sigma_t^2}(s_1^2, \ldots, s_t^2)$ is

$$\frac{1}{\sigma_t^2}(s_1^2, \ldots, s_t^2) \sim \text{Gamma}(\hat{\alpha}_t, \hat{\beta}_t) \tag{3.21}$$

with $\hat{\alpha}_t$, $\hat{\beta}_t$ determined by the recursive equations (3.20). As with the
adaptive filter for the process mean, a box and whisker control charting technique can be considered, along with the more conventional Shewhart-type plotting of $m_{t}^{-1}$ with fixed control limits. Note that $m_{t}^{-1}$ is an adaptive exponentially weighted moving average of sample variances. On our box and whisker plots, an '0' near the middle of the box represents $m_{t}^{-1}$, the estimate of $\sigma_{t}^{2}$. An "S" on the chart at time $t$ represents the observed sample variance at time $t$. The bottom whisker corresponds to a lower percentile of the estimated posterior distribution of $\sigma_{t}^{2}$, which can be obtained using (3.21). When the bottom whisker is larger than the target variance, the process variance is considered out-of-control. The bottom of the box corresponds to a less extreme percentile of the distribution. An illustration is given in Figure 3.5.

![Figure 3.5](image)

Figure 3.5. Box plots for a process variance

The first box-plot in Figure 3.5 represents an out-of-control signal since the lower whisker is greater than $\sigma_{0}^{2}$, the target variance. The second box-plot represents a marginally in-control or warning signal.
Figure 3.6 illustrates the adaptive algorithm at work in box-plot form, using simulated $s_t^2$ values, for $t = 1, 2, \ldots, 10$.

The Shewhart-type control limits consist of constant boundaries. When the estimate $\hat{\sigma}_t^2 = \hat{m}_t^{-1}$ falls outside these boundaries, the process variance is deemed out of control.

Again, simulations using an i.i.d. Gamma model for the sample variances were performed to obtain empirical estimates of the ARLs for both of the control procedures. Tables (3.10-3.13) and (3.14-3.17) contain these ARLs for the schemes based on samples of size $n = 2, 6, 10, \text{ and } 20$. Again, standard errors of the simulation estimates were roughly 10% of the estimated means. In Tables (3.10-3.13) the Shewhart control limits correspond to the $80^{th}$, $85^{th}$, $90^{th}$, $95^{th}$, $97.5^{th}$, $99^{th}$, and $99.9^{th}$ percentiles of the distribution of $s_t^2$ under an i.i.d. in-control model, where the in-control variance is taken to be $\sigma^2 = 1$. In Tables (3.14-3.17) the control limit quantiles refer to the quantiles of the Gamma distribution (3.21) used to construct the box-plot whiskers. In both sets of tables, the heading $\sigma^2$ refers to the true variance under the i.i.d. model.

To start the algorithm (3.20), $\alpha_0$ and $\beta_0$ must be specified. To obtain appropriate starting values for $\alpha_0$ and $\beta_0$, a trial run of length 100 was performed using sample variances generated from the i.i.d. in-control Gamma model. The trial run was started using $\alpha_0 = \beta_0 = 1$. The sample means of the 100 $\alpha_t$'s and $\beta_t$'s generated in the trial run were used as starting values for $\alpha_0$ and $\beta_0$, respectively, in the larger
0 - Adaptive estimate of process variance
S - Sample variance
Box and Whiskers extend to the 20th and .5th percentiles, respectively, of the estimated posterior distribution of the process variance.

Figure 3.6. Adaptive filter for process variance
### Table 3.10. ARLs: Adaptive filter for process variance, Shewhart control limits, n=2

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>1.64</th>
<th>2.07</th>
<th>2.71</th>
<th>3.84</th>
<th>5.02</th>
<th>6.64</th>
<th>7.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>9.54</td>
<td>13.24</td>
<td>24.85</td>
<td>67.84</td>
<td>180.43</td>
<td>670.90</td>
<td>1836.02</td>
</tr>
<tr>
<td>1.50</td>
<td>4.61</td>
<td>6.76</td>
<td>9.30</td>
<td>18.95</td>
<td>34.05</td>
<td>94.94</td>
<td>240.65</td>
</tr>
<tr>
<td>1.75</td>
<td>4.24</td>
<td>6.05</td>
<td>8.78</td>
<td>14.94</td>
<td>26.21</td>
<td>64.47</td>
<td>101.62</td>
</tr>
<tr>
<td>2.00</td>
<td>3.87</td>
<td>4.52</td>
<td>6.37</td>
<td>13.01</td>
<td>25.09</td>
<td>42.90</td>
<td>64.30</td>
</tr>
<tr>
<td>3.00</td>
<td>2.56</td>
<td>3.22</td>
<td>4.24</td>
<td>6.04</td>
<td>9.02</td>
<td>14.75</td>
<td>24.82</td>
</tr>
<tr>
<td>4.00</td>
<td>2.28</td>
<td>2.82</td>
<td>2.97</td>
<td>4.74</td>
<td>5.94</td>
<td>7.98</td>
<td>13.04</td>
</tr>
<tr>
<td>5.00</td>
<td>2.09</td>
<td>2.77</td>
<td>2.65</td>
<td>3.31</td>
<td>3.95</td>
<td>5.87</td>
<td>10.49</td>
</tr>
</tbody>
</table>

### Table 3.11. ARLs: Adaptive filter for process variance, Shewhart control limits, n=6

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>1.37</th>
<th>1.62</th>
<th>1.85</th>
<th>2.21</th>
<th>2.57</th>
<th>3.02</th>
<th>3.35</th>
<th>4.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>7.37</td>
<td>10.11</td>
<td>18.31</td>
<td>35.46</td>
<td>82.32</td>
<td>260.48</td>
<td>596.01</td>
<td>3927.96</td>
</tr>
<tr>
<td>1.25</td>
<td>4.31</td>
<td>5.42</td>
<td>7.97</td>
<td>14.40</td>
<td>26.11</td>
<td>54.14</td>
<td>117.99</td>
<td>550.77</td>
</tr>
<tr>
<td>1.50</td>
<td>3.05</td>
<td>2.96</td>
<td>5.06</td>
<td>7.57</td>
<td>13.16</td>
<td>20.84</td>
<td>26.93</td>
<td>146.92</td>
</tr>
<tr>
<td>1.75</td>
<td>2.17</td>
<td>2.54</td>
<td>3.05</td>
<td>4.93</td>
<td>8.06</td>
<td>14.71</td>
<td>19.34</td>
<td>53.39</td>
</tr>
<tr>
<td>2.00</td>
<td>2.18</td>
<td>2.22</td>
<td>2.57</td>
<td>3.33</td>
<td>5.15</td>
<td>7.20</td>
<td>12.13</td>
<td>29.25</td>
</tr>
<tr>
<td>3.00</td>
<td>1.42</td>
<td>1.51</td>
<td>1.52</td>
<td>2.20</td>
<td>2.37</td>
<td>2.62</td>
<td>3.58</td>
<td>6.71</td>
</tr>
<tr>
<td>4.00</td>
<td>1.21</td>
<td>1.24</td>
<td>1.45</td>
<td>1.49</td>
<td>1.61</td>
<td>1.79</td>
<td>2.30</td>
<td>3.99</td>
</tr>
<tr>
<td>5.00</td>
<td>1.11</td>
<td>1.21</td>
<td>1.28</td>
<td>1.31</td>
<td>1.36</td>
<td>1.59</td>
<td>1.58</td>
<td>2.68</td>
</tr>
</tbody>
</table>
Table 3.12. ARLs: Adaptive filter for process variance, Shewhart control limits, n=10

<table>
<thead>
<tr>
<th>Control limit</th>
<th>1.36</th>
<th>1.48</th>
<th>1.63</th>
<th>1.88</th>
<th>2.11</th>
<th>2.41</th>
<th>2.62</th>
<th>3.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>6.79</td>
<td>9.01</td>
<td>14.15</td>
<td>31.06</td>
<td>60.97</td>
<td>195.50</td>
<td>382.86</td>
<td>2455.45</td>
</tr>
<tr>
<td>1.25</td>
<td>3.52</td>
<td>5.06</td>
<td>6.25</td>
<td>8.94</td>
<td>17.83</td>
<td>39.66</td>
<td>73.85</td>
<td>317.25</td>
</tr>
<tr>
<td>1.50</td>
<td>2.05</td>
<td>2.85</td>
<td>3.69</td>
<td>4.90</td>
<td>7.29</td>
<td>15.11</td>
<td>18.67</td>
<td>53.79</td>
</tr>
<tr>
<td>1.75</td>
<td>1.80</td>
<td>2.09</td>
<td>2.30</td>
<td>3.11</td>
<td>4.08</td>
<td>7.42</td>
<td>8.79</td>
<td>18.96</td>
</tr>
<tr>
<td>2.00</td>
<td>1.51</td>
<td>1.61</td>
<td>1.79</td>
<td>2.37</td>
<td>3.66</td>
<td>4.72</td>
<td>5.39</td>
<td>10.12</td>
</tr>
<tr>
<td>3.00</td>
<td>1.09</td>
<td>1.14</td>
<td>1.43</td>
<td>1.32</td>
<td>1.48</td>
<td>1.95</td>
<td>2.00</td>
<td>3.18</td>
</tr>
<tr>
<td>4.00</td>
<td>1.07</td>
<td>1.06</td>
<td>1.09</td>
<td>1.15</td>
<td>1.21</td>
<td>1.40</td>
<td>1.58</td>
<td>1.88</td>
</tr>
<tr>
<td>5.00</td>
<td>1.02</td>
<td>1.04</td>
<td>1.04</td>
<td>1.11</td>
<td>1.09</td>
<td>1.18</td>
<td>1.19</td>
<td>1.39</td>
</tr>
</tbody>
</table>

Table 3.13. ARLs: Adaptive filter for process variance, Shewhart control limits, n=20

<table>
<thead>
<tr>
<th>Control limit</th>
<th>1.26</th>
<th>1.33</th>
<th>1.43</th>
<th>1.59</th>
<th>1.73</th>
<th>1.91</th>
<th>2.03</th>
<th>2.31</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>6.23</td>
<td>8.19</td>
<td>13.35</td>
<td>29.62</td>
<td>47.03</td>
<td>171.58</td>
<td>321.54</td>
<td>1814.69</td>
</tr>
<tr>
<td>1.25</td>
<td>2.72</td>
<td>2.73</td>
<td>3.67</td>
<td>6.15</td>
<td>11.00</td>
<td>16.24</td>
<td>34.33</td>
<td>124.33</td>
</tr>
<tr>
<td>1.50</td>
<td>1.59</td>
<td>1.66</td>
<td>2.25</td>
<td>2.65</td>
<td>3.80</td>
<td>5.81</td>
<td>8.32</td>
<td>20.34</td>
</tr>
<tr>
<td>1.75</td>
<td>1.23</td>
<td>1.44</td>
<td>1.51</td>
<td>1.68</td>
<td>2.50</td>
<td>3.49</td>
<td>3.96</td>
<td>6.79</td>
</tr>
<tr>
<td>2.00</td>
<td>1.14</td>
<td>1.21</td>
<td>1.24</td>
<td>1.39</td>
<td>1.96</td>
<td>2.62</td>
<td>2.55</td>
<td>4.27</td>
</tr>
<tr>
<td>3.00</td>
<td>1.02</td>
<td>1.04</td>
<td>1.08</td>
<td>1.10</td>
<td>1.08</td>
<td>1.14</td>
<td>1.23</td>
<td>1.31</td>
</tr>
<tr>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.02</td>
<td>1.00</td>
<td>1.01</td>
<td>1.10</td>
<td>1.15</td>
</tr>
<tr>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.02</td>
</tr>
</tbody>
</table>
### Table 3.14. ARLs: Adaptive filter for process variance, box plot type gamma percentile limits, n=2

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>.80</th>
<th>.85</th>
<th>.90</th>
<th>.95</th>
<th>.975</th>
<th>.99</th>
<th>.995</th>
<th>.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>8.67</td>
<td>11.58</td>
<td>15.68</td>
<td>27.38</td>
<td>50.57</td>
<td>127.04</td>
<td>213.94</td>
<td>1109.44</td>
</tr>
<tr>
<td>1.25</td>
<td>5.47</td>
<td>8.28</td>
<td>9.20</td>
<td>14.33</td>
<td>27.25</td>
<td>47.62</td>
<td>78.75</td>
<td>269.63</td>
</tr>
<tr>
<td>1.50</td>
<td>4.75</td>
<td>5.50</td>
<td>8.77</td>
<td>12.01</td>
<td>18.69</td>
<td>29.35</td>
<td>40.91</td>
<td>116.76</td>
</tr>
<tr>
<td>1.75</td>
<td>3.90</td>
<td>4.71</td>
<td>6.83</td>
<td>8.81</td>
<td>13.85</td>
<td>17.69</td>
<td>24.94</td>
<td>67.77</td>
</tr>
<tr>
<td>2.00</td>
<td>3.58</td>
<td>4.25</td>
<td>5.92</td>
<td>7.76</td>
<td>8.73</td>
<td>15.25</td>
<td>20.63</td>
<td>37.46</td>
</tr>
<tr>
<td>3.00</td>
<td>3.02</td>
<td>2.87</td>
<td>3.46</td>
<td>3.99</td>
<td>5.40</td>
<td>6.44</td>
<td>8.63</td>
<td>14.09</td>
</tr>
<tr>
<td>4.00</td>
<td>2.21</td>
<td>2.31</td>
<td>2.62</td>
<td>3.51</td>
<td>4.30</td>
<td>4.58</td>
<td>5.75</td>
<td>8.06</td>
</tr>
<tr>
<td>5.00</td>
<td>1.98</td>
<td>2.22</td>
<td>2.54</td>
<td>3.12</td>
<td>3.30</td>
<td>3.72</td>
<td>4.45</td>
<td>6.67</td>
</tr>
</tbody>
</table>

### Table 3.15. ARLs: Adaptive filter for process variance, box plot type gamma percentile limits, n=6

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>.80</th>
<th>.85</th>
<th>.90</th>
<th>.95</th>
<th>.975</th>
<th>.99</th>
<th>.995</th>
<th>.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>6.13</td>
<td>7.27</td>
<td>11.68</td>
<td>20.37</td>
<td>38.42</td>
<td>94.98</td>
<td>196.05</td>
<td>603.26</td>
</tr>
<tr>
<td>1.25</td>
<td>3.98</td>
<td>5.43</td>
<td>6.37</td>
<td>7.62</td>
<td>16.73</td>
<td>26.23</td>
<td>38.45</td>
<td>118.09</td>
</tr>
<tr>
<td>1.50</td>
<td>2.50</td>
<td>3.43</td>
<td>3.82</td>
<td>5.17</td>
<td>8.01</td>
<td>14.12</td>
<td>18.44</td>
<td>41.53</td>
</tr>
<tr>
<td>1.75</td>
<td>2.16</td>
<td>2.17</td>
<td>2.72</td>
<td>3.27</td>
<td>4.00</td>
<td>6.74</td>
<td>8.69</td>
<td>14.70</td>
</tr>
<tr>
<td>2.00</td>
<td>1.76</td>
<td>1.88</td>
<td>2.48</td>
<td>2.86</td>
<td>3.51</td>
<td>4.46</td>
<td>6.69</td>
<td>8.71</td>
</tr>
<tr>
<td>3.00</td>
<td>1.33</td>
<td>1.39</td>
<td>1.52</td>
<td>1.68</td>
<td>2.07</td>
<td>2.34</td>
<td>2.50</td>
<td>3.19</td>
</tr>
<tr>
<td>4.00</td>
<td>1.25</td>
<td>1.15</td>
<td>1.30</td>
<td>1.33</td>
<td>1.44</td>
<td>1.67</td>
<td>1.71</td>
<td>2.00</td>
</tr>
<tr>
<td>5.00</td>
<td>1.12</td>
<td>1.09</td>
<td>1.20</td>
<td>1.35</td>
<td>1.40</td>
<td>1.59</td>
<td>1.58</td>
<td>1.84</td>
</tr>
</tbody>
</table>
Table 3.16. ARLs: Adaptive filter for process variance, box plot type gamma percentile limits, n=10

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>0.80</th>
<th>0.85</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.99</th>
<th>0.995</th>
<th>0.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>6.26</td>
<td>7.18</td>
<td>11.96</td>
<td>23.80</td>
<td>35.60</td>
<td>105.07</td>
<td>202.65</td>
<td>638.58</td>
</tr>
<tr>
<td>1.25</td>
<td>3.21</td>
<td>3.34</td>
<td>5.02</td>
<td>5.98</td>
<td>10.75</td>
<td>19.21</td>
<td>30.16</td>
<td>95.94</td>
</tr>
<tr>
<td>1.50</td>
<td>1.99</td>
<td>2.71</td>
<td>2.56</td>
<td>3.29</td>
<td>5.18</td>
<td>9.79</td>
<td>10.50</td>
<td>23.78</td>
</tr>
<tr>
<td>1.75</td>
<td>1.59</td>
<td>1.96</td>
<td>2.04</td>
<td>2.63</td>
<td>3.36</td>
<td>4.51</td>
<td>5.36</td>
<td>10.56</td>
</tr>
<tr>
<td>2.00</td>
<td>1.42</td>
<td>1.49</td>
<td>1.73</td>
<td>2.00</td>
<td>2.25</td>
<td>3.23</td>
<td>3.03</td>
<td>6.05</td>
</tr>
<tr>
<td>3.00</td>
<td>1.17</td>
<td>1.12</td>
<td>1.20</td>
<td>1.28</td>
<td>1.33</td>
<td>1.62</td>
<td>1.53</td>
<td>2.10</td>
</tr>
<tr>
<td>4.00</td>
<td>1.05</td>
<td>1.03</td>
<td>1.14</td>
<td>1.18</td>
<td>1.22</td>
<td>1.23</td>
<td>1.42</td>
<td>1.64</td>
</tr>
<tr>
<td>5.00</td>
<td>1.04</td>
<td>1.02</td>
<td>1.03</td>
<td>1.07</td>
<td>1.11</td>
<td>1.14</td>
<td>1.12</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Table 3.17. ARLs: Adaptive filter for process variance, box plot type gamma percentile limits, n=20

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>0.80</th>
<th>0.85</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
<th>0.99</th>
<th>0.995</th>
<th>0.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>5.59</td>
<td>7.53</td>
<td>12.33</td>
<td>22.85</td>
<td>43.58</td>
<td>113.58</td>
<td>210.31</td>
<td>650.12</td>
</tr>
<tr>
<td>1.25</td>
<td>2.38</td>
<td>2.47</td>
<td>3.41</td>
<td>5.20</td>
<td>8.02</td>
<td>13.72</td>
<td>23.45</td>
<td>56.94</td>
</tr>
<tr>
<td>1.50</td>
<td>1.62</td>
<td>1.87</td>
<td>2.07</td>
<td>2.90</td>
<td>3.51</td>
<td>5.24</td>
<td>7.36</td>
<td>11.29</td>
</tr>
<tr>
<td>1.75</td>
<td>1.22</td>
<td>1.30</td>
<td>1.51</td>
<td>1.70</td>
<td>2.19</td>
<td>2.58</td>
<td>3.52</td>
<td>5.96</td>
</tr>
<tr>
<td>2.00</td>
<td>1.09</td>
<td>1.19</td>
<td>1.30</td>
<td>1.40</td>
<td>1.58</td>
<td>1.87</td>
<td>1.99</td>
<td>3.16</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.02</td>
<td>1.05</td>
<td>1.04</td>
<td>1.12</td>
<td>1.12</td>
<td>1.16</td>
<td>1.29</td>
</tr>
<tr>
<td>4.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
<td>1.05</td>
<td>1.04</td>
<td>1.12</td>
</tr>
<tr>
<td>5.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.01</td>
<td>1.02</td>
</tr>
</tbody>
</table>
simulation studies. Table 3.18 gives the recommended starting values for $\alpha_0$ and $\beta_0$ for sample sizes $n=2, 6, 10, 20$. In each case, $m_0$ was taken to be $\alpha_0/\beta_0$.

The two control procedures described above can be evaluated by comparing out-of-control ARLs given the same in-control ARLs. For each of the sample sizes $n=2, 6, 10, 20$, the preferred control procedure depends on the type of sensitivity desired. For very large percent increases (300–500%) in the process variance relative to the in-control process variance, the Shewhart limit adaptive control procedure appears to signal faster than the box-plot procedure. If the percent increase is not extremely large, the box-plot procedure generally performs better. In practice, a small percent increase in process variance might be cause for concern. In this case, the box-plot procedure is preferable to the Shewhart procedure for all sample sizes. Tables (3.10–3.13) and (3.14–3.17) can be used to determine at what percent increase in the process variance the Shewhart control limits appear to become preferable.

Table 3.18. Starting values for adaptive control procedures for process variance

<table>
<thead>
<tr>
<th>n</th>
<th>$\alpha_0$</th>
<th>$\beta_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.9</td>
<td>6.7</td>
</tr>
<tr>
<td>6</td>
<td>11.1</td>
<td>11.2</td>
</tr>
<tr>
<td>10</td>
<td>17.7</td>
<td>18.2</td>
</tr>
<tr>
<td>20</td>
<td>37.3</td>
<td>37.2</td>
</tr>
</tbody>
</table>
Both of the approaches described above were compared to the standard Shewhart $s^2$ control chart, under an i.i.d. model. Table 3.19 gives the ARL values of the $s^2$ procedure, for sample size n=6 and can be compared to Tables 3.11 and 3.15. Both adaptive control procedures appear to be quicker to detect small increases in the process variance. For larger increases in the process variance, however, the ARLs are similar, with the adaptive control procedures showing no improvement over the Shewhart $s^2$ charting procedure.

To construct adaptive control charts for monitoring a process variance, the procedure outlined in Chapter II is again recommended.

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>1.46</th>
<th>1.62</th>
<th>1.85</th>
<th>2.21</th>
<th>2.57</th>
<th>3.02</th>
<th>3.35</th>
<th>4.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>5.00</td>
<td>6.67</td>
<td>10.00</td>
<td>20.02</td>
<td>40.09</td>
<td>100.04</td>
<td>201.98</td>
<td>1106.31</td>
</tr>
<tr>
<td>1.25</td>
<td>3.09</td>
<td>3.83</td>
<td>5.18</td>
<td>8.71</td>
<td>14.72</td>
<td>29.55</td>
<td>50.42</td>
<td>187.79</td>
</tr>
<tr>
<td>1.50</td>
<td>2.31</td>
<td>2.72</td>
<td>3.44</td>
<td>5.16</td>
<td>7.81</td>
<td>13.59</td>
<td>20.82</td>
<td>59.87</td>
</tr>
<tr>
<td>1.75</td>
<td>1.90</td>
<td>2.17</td>
<td>2.61</td>
<td>3.63</td>
<td>5.08</td>
<td>7.99</td>
<td>11.35</td>
<td>27.18</td>
</tr>
<tr>
<td>2.00</td>
<td>1.66</td>
<td>1.85</td>
<td>2.15</td>
<td>2.83</td>
<td>3.74</td>
<td>5.46</td>
<td>7.32</td>
<td>15.32</td>
</tr>
<tr>
<td>3.00</td>
<td>1.27</td>
<td>1.34</td>
<td>1.45</td>
<td>1.68</td>
<td>1.96</td>
<td>2.43</td>
<td>2.87</td>
<td>4.41</td>
</tr>
<tr>
<td>4.00</td>
<td>1.15</td>
<td>1.18</td>
<td>1.24</td>
<td>1.36</td>
<td>1.50</td>
<td>1.72</td>
<td>1.91</td>
<td>2.54</td>
</tr>
<tr>
<td>5.00</td>
<td>1.09</td>
<td>1.11</td>
<td>1.15</td>
<td>1.22</td>
<td>1.31</td>
<td>1.43</td>
<td>1.55</td>
<td>1.89</td>
</tr>
</tbody>
</table>
IV. A UNIVARIATE PROCESS CONTROL MODEL

A. The General Linear Stochastic Control Model

The general model in Stochastic Control Theory, as described by Astrom (1970) consists of the matrix linear system

\[ Y_t = F_t \theta_t + \epsilon_t \quad t = 1, 2, \ldots \quad (4.1) \]

\[ \theta_t = G_t \theta_{t-1} + \Gamma_t u_{t-1} + \nu_t \]

where \( F_t, G_t, \) and \( \Gamma_t \) are known. The data at time \( t \) are represented by \( Y_t \), which may be either scalar or vector valued. \( \theta_t \) is an \( nx1 \) state vector, \( u_t \) a \( px1 \) vector of control variables which depends on the data through time \( t \), and \( \{\epsilon_t, t = 1, 2, \ldots\} \) and \( \{\nu_t, t = 1, 2, \ldots\} \) are independent sequences of independent multivariate normal random variables with zero mean values and covariances

\[ \text{Cov}(\epsilon_t, \epsilon_t) = V_1 \quad (4.2) \]

\[ \text{Cov}(\nu_t, \nu_t) = V_2 \]

The matrices \( V_1 \) and \( V_2 \) may also depend on \( t \). It is assumed that the initial state \( \theta_0 \) is multivariate normal with

\[ E(\theta_0) = \mu \]
and \[ \text{Cov}(\theta_0, \theta_0) = \mathbf{V}_0 \].

The standard characterization of the performance of the system (4.1) is the scalar risk function

\[
J(n) = \mathbb{E} \left\{ \sum_{t=1}^{n} (\theta_t'Q_1\theta_t + u_{t-1}'Q_2u_{t-1}) \right\}.
\]

(4.3)

The matrix \( Q_1 \) here is symmetric and nonnegative definite and the matrix \( Q_2 \) is assumed to be positive definite. These matrices may also depend on \( t \).

The usual stochastic control problem is then to find a control strategy (i.e., a sequence of inputs \( \{u_t\} \)) for the system described by (4.1, 4.2) such that the criterion (4.3) is minimal.

B. A Special Case of the General Linear Stochastic Control Problem and a New Performance Criterion

Here we will consider the special case of the general model (4.1) consisting of the scalar linear system with, respectively, the observation and system equations

\[
y_t = \theta_t + \varepsilon_t \quad t = 1, 2, \ldots
\]

(4.4)

\[
\theta_t = \theta_{t-1} + u_{t-1} + \nu_t
\]
with error structure as in (2.2). This is the steady model (2.1) with the addition of the control variable \( u_{t-1} \). Note that if no control actions are taken, \( u_t = 0 \), and the process means \( \theta_t \) evolve as a random walk.

Let \( y^t \) represent all observed data through time \( t \). To initialize the model given in (4.4) we will assume that conditional on \( y^0 \) (the past history of the \( y_t \) process), \( \theta_0 \) is normally distributed with

\[
E(\theta_0 | y^0) = \hat{\theta}_0
\]

and

\[
V(\theta_0 | y^0) = q .
\]

We will take \( q \) to be the steady state variance of \( \theta_t | y^t \), and have from Appendix A that

\[
q = (\sigma_y^2 + (\sigma_\epsilon^2 + 4\sigma_y^2\sigma_\epsilon^2)^{1/2})/2 .
\]

(This choice of \( q \) is appropriate if \( y^0 \) represents a substantial past history.)

Before any of \( y_1, y_2, \ldots \) are observed, \( \hat{\theta}_0 \) is our best estimate of \( \theta_0 \). From (4.4), we have that

\[
y_1 | \theta_1 \sim N(\theta_1, \sigma_\epsilon^2) \quad (4.5)
\]
and \[ \theta \mid y \sim N(\hat{\theta}_0 + u_0, q + \sigma_y^2), \]

where \( u_0 \) represents a first control action and is a function of \( y^0 \). Applying Bayes' theorem, we have that

\[ \theta \mid y \sim N(\hat{\theta}, q), \]  \hspace{1cm} (4.6)

where \( \hat{\theta} = \hat{\theta}_0 + u_0 + k(y_1 - (\hat{\theta}_0 + u_0)) \)

and \[ k = \frac{(q + \sigma_y^2)}{\sigma_y^2 + q + \sigma_y^2}. \]

In general,

\[ y_t \mid \theta_t \sim N(\theta_t, \sigma_y^2) \]

and \[ \theta \mid y^{t-1} \sim N(\hat{\theta}_{t-1} + u_{t-1}, p), \]

where \( p = q + \sigma_y^2 \).

Again, by Bayes' theorem,

\[ \theta \mid y \sim N(\hat{\theta}_t, q), \]

where \[ \hat{\theta}_t = \hat{\theta}_t + u_{t-1} + \left( \frac{p}{\sigma_y^2 + p} \right) (y_t - (\hat{\theta}_{t-1} + u_{t-1})). \]  \hspace{1cm} (4.7)
Equation (4.7) gives a generalized Kalman filter estimate of $\theta_t$ given the data through time $t$. And using (4.7) and substituting for $p$,

$$\hat{\theta}_t | y^{t-1} \sim N(\hat{\theta}_{t-1} + u_{t-1}, \sigma^2_v)$$

(4.8)

where $\sigma^2_v$ is the system variance.

The usual characterization of the performance of the system (4.4) would be the special case of the expected loss function (4.3)

$$J(n) = E \left\{ \sum_{t=1}^{n} (k_1 \theta_t^2 + k_2 u_t^2) \right\}$$

(4.9)

$$= k_1 E \left\{ \sum_{t=1}^{n} \theta_t^2 \right\} + k_2 E \left\{ \sum_{t=1}^{n} u_t^2 \right\}$$

where $c = k_2 / k_1$, $k_1$ represents the cost per squared unit deviation from a target of 0, and $k_2 u_t^2$ represents the cost associated with the control action or input $u_t$. Notice that for purposes of deriving an optimal control strategy, we can replace (4.9) with the performance criterion

$$E \left\{ \sum_{t=1}^{n} (\theta_t^2 + cu_t^2) \right\}$$

(4.10)

and the standard optimal control problem is to find a control strategy for the system described by (4.4) such that the criterion (4.10) is minimal.
Astrom (1970) solves the general form of this problem using dynamic programming. Specialization of his result shows that the criterion (4.10) is minimized by taking

$$u_t = -D_t \hat{\theta}_t,$$

where $$\hat{\theta}_t = E(\theta_t | y_t)$$, the generalized Kalman filter estimate of $$\theta_t$$ given the data through time $$t$$, and $$D_t$$ is a feedback constant at time $$t$$. The feedback constants $$D_t$$ depend only on $$t$$ and $$c$$, the parameter of the loss function, and can be calculated before any data is observed. For this special case,

$$D_t = (a_{t+1}/(c + a_{t+1})$$, \hspace{1cm} (4.11) \tag{4.11}$$

where $$a_t = (a_{t+1}(c+1) + c)/(a_{t+1} + c)$$, \hspace{1cm} (4.12) \tag{4.12}$$

and $$a_\infty = 1$$.

The solution to this control problem is a time varying one in that it depends on the time, $$n$$, at which control is to be terminated. In general, one may be interested in the steady state optimal feedback control action, where the period over which the control is to be applied is effectively infinite. For fixed $$t$$, as $$n$$ tends to infinity, the value of $$a_t$$ derived from the recursive equation (4.12) tends to a constant $$a_\infty$$, and $$D_t$$ tends to a constant.
\[ D_\infty = a_\infty / (c + a_\infty), \]

which can be determined from (4.11) and (4.12). This steady state solution, along with the generalized Kalman filter estimate \( \hat{\theta}_t \), provide the "infinite horizon" optimal control action

\[ u_t = - D_\infty \hat{\theta}_t. \]

In the present case, it can be shown that

\[ D_\infty = (1 + (1+4c)^2) / (1+2c + (1+4c)^2). \]

For the case \( c=0 \), implying no restriction on the magnitude of \( u_t \), \( D_\infty = 1 \) and \( u_t = - \hat{\theta}_t \). For \( c \) very large, implying a severe penalty for large \( u_t^2 \), \( D_\infty = 0 \), and \( u_t = 0 \).

In the control theory literature, performance criteria of the type (4.3) are generally used. Although numerical solutions can, at least theoretically, be obtained for any system performance criterion, analytical solutions in closed form appear in the literature only for such quadratic criteria.

The solution of the special quadratic loss problem above is typical in that it calls for adjustment of the process at each time period. In the usual applications cited for control theory, the action \( u_t \) is made by computer or automatic controller and the cost \( u_t^2 \) is included in (4.10) to restrict the magnitude of the input \( u_t \).
In contrast to the usual form of "optimal" control strategies, standard statistical quality control philosophy calls for adjustment of a process only when a clearly extreme observation is encountered, i.e., there is strong evidence of misadjustment. In cases where adjustment is to be made manually, the cost involved is typically a setup (or fixed) cost that does not depend on the magnitude of the adjustment. To reflect this philosophy, here we propose to use a different performance criterion.

Consider characterizing the performance of the system described by (4.4) by the expected loss

\[ J'(n) = E \left\{ \frac{1}{n} \sum_{t=1}^{n} (k_1 \theta_t^2 + k_2 \delta(u_{t-1})) \right\} \]

\[ = k_1 E \left\{ \frac{1}{n} \sum_{t=1}^{n} (\theta_t^2 + c\delta(u_{t-1})) \right\}, \]

where
\[ \delta(a) = \begin{cases} 0 & \text{if } a = 0 \\ 1 & \text{if } a \neq 0 \end{cases} \]

and \[ c = \frac{k_2}{k_1}. \]

As before, \( k_1 \theta_t^2 \) represents the cost associated with a deviation of \(|\theta_t|\) from target. But now \( k_2 \) represents the (fixed) cost of taking a control action of any magnitude.
Clearly, optimization of \( J'(n) \) is equivalent to optimization of

\[
L(n) = E \left\{ \sum_{t=1}^{n} (\theta_t^2 + c\delta(u_{t-1})) \right\} .
\] (4.13)

The problem in the remainder of this chapter is to find a control strategy for the system described by (4.4) such that the new criterion (4.13) is minimal. [Box and Jenkins (1963) and Bather (1963) have considered related problems in which a fixed cost is associated with taking a control action, and have results which apply to the infinite horizon control problem.]

C. A Functional Equation

Define

\[
R_1(\theta) = \min_{v} \int (x^2 + c\delta(v))h(x;v+\theta)dx ,
\] (4.14)

where \( h(x;u) \) is the \( N(u,p) \) density.

Then

\[
R_1(\theta) = \min_{v} \{ p + (\theta+v)^2 + c\delta(v) \}
\]

\[= p + \min(c, \theta^2) .\]

Note that

\[
\min L(1) = \min_{u_0} E\{\theta_1^2 + c\delta(u_0)\}
\]
\[
\begin{align*}
\theta_0 &= \min \{ p + (\theta_0 + u_0)^2 + c\delta(u_0) \} \\
&= p + \min(c, \theta_0^2) \\
&= R_1(\theta_0) ,
\end{align*}
\]

and that \( u_0 \) minimizing \( L(1) \) is

\[
\begin{align*}
\theta_0 &= -\hat{\theta}_0 \quad \text{if} \quad |\hat{\theta}_0| \geq c^{\frac{1}{2}} \\
\theta_0 &= 0 \quad \quad \quad \quad \quad \quad \text{if} \quad |\hat{\theta}_0| < c^{\frac{1}{2}} .
\end{align*}
\]

Now, for \( \lambda > 1 \) define recursively,

\[
R_\lambda(\theta) = \min \{ p + (\theta + v)^2 + c\delta(v) + \int R_{\lambda-1}(x)f(x; \theta + v)dx \}, \tag{4.15}
\]

where \( f(x; \mu) \) is the \( N(\mu, \sigma^2_v) \) density. From (4.8),

\[
\hat{\theta}_1 \mid y^0 \sim N(\hat{\theta}_0 + u_0, \sigma^2_v) ,
\]

so that \( R_\lambda(\theta) \) is the minimum of a quantity consisting of the expected loss over \( u_0 \) and \( \theta_1 \) plus the expected value of the conditional optimum risk suffered thereafter, given \( \hat{\theta}_1 \). Note that
If we further define \( v_\chi(\theta) \) to be a value of \( v \) associated with \( R_\chi(\theta) \), then optimal \( u_t \)'s can be identified as

\[
u_{n-k} = v_\chi(\hat{\theta}_{n-k}) \quad k = 1, 2, \ldots, n.
\]

The \( u_t \)'s can be determined by solving the functional equation (4.15).

D. Solution of the Functional Equation

The initial condition for the functional equation (4.15) is from (4.14),

\[
R_1(\theta) = p + \min(c, \theta^2),
\]

with \( v_1(\theta) = -\theta \) if \( |\theta| \geq c^{\frac{1}{2}} \)

\[
= 0 \quad \text{if} \quad |\theta| < c^{\frac{1}{2}}.
\]

Note that the critical constant defining the optimal control action, \( c^{\frac{1}{2}} \), is the positive solution to the equation

\[
g_1(\theta) \equiv \theta^2 - c = 0.
\]

Let \( k_1 \) stand for \( c^{\frac{1}{2}} \).
Next, using (4.14) and (4.15), notice that

\[ R_2(\theta) = \min_v \{ p + (\theta + v)^2 + c\delta(v) + \int R_1(x)f(x;v+\theta)dx \} \]

\[ = \min_v \{ 2p + (\theta + v)^2 + c\delta(v) + \int \min(c,x^2)f(x;v+\theta)dx \} \]

\[ = \min_v \{ 2p + c + (\theta + v)^2 + c\delta(v) + \int_{[x^2 < c]} (x^2 - c)f(x;v+\theta)dx \} \]

\[ = 2p + c + \min_v \{ \theta^2 + c\delta(v) + \int_{[x^2 < c]} (x^2 - c)f(x;v+\theta)dx \} \]

By Appendix C, the integral is minimized for \( v = -\theta \). Thus,

\[ R_2(\theta) = 2p + c + \min(c + \int_{[x^2 < c]} (x^2 - c)f(x;0)dx) \]

\[ \theta^2 + \int_{[x^2 < c]} (x^2 - c)f(x;\theta)dx \]

and \( \nu_2(\theta) = -\theta \) if \(|\theta| \geq k_2\)

\[ = 0 \quad \text{if} \quad |\theta| < k_2 \]

where \( k_2 \) is the positive solution to the equation
\[ g_2(\theta) = (\theta^2-c) + \int_{[x^2<c]} (x^2-c)[f(x;\theta) - f(x;0)]dx \]
\[ = (\theta^2-c) + \int_{-k_1}^{k_1} g_1(x)[f(x;\theta) - f(x;0)]dx \]
\[ = 0. \]

(Note that \( g_2(\theta) \) is symmetric in \( \theta \) about zero, and by Lemma 4.1 of Appendix C is increasing in \( |\theta| \).)

We will now argue that in general the functional equation has a solution which is of the form

\[ R_{\lambda}(\theta) = c_{\lambda} + \min(w_{\lambda}, h_{\lambda}(\theta)), \quad (4.16) \]

where \( h_{\lambda}(\theta) \) is symmetric about zero, and increasing in \( |\theta| \) and \( c_{\lambda} \) and \( w_{\lambda} \) are constants. This is clearly true for \( \lambda = 1 \) and \( \lambda = 2 \). Proceeding by induction, we assume that (4.16) holds for \( \lambda \) and we will then show that it holds for \( \lambda + 1 \).

From (4.15),

\[ R_{\lambda+1}(\theta) = \min_{v} \{ p + (\theta+v)^2 + c\delta(v) + \int R_{\lambda}(x)f(x;v+\theta)dx \} \]
\[ = \min_{v} \{ p+(\theta+v)^2+c\delta(v) + \int (c_{\lambda}+\min(w_{\lambda}, h_{\lambda}(x)))f(x;v+\theta)dx \} \]
\[= \min_{\nu} \left( p + c_\nu + (\theta + \nu)^2 + c\delta(\nu) + w_\nu + \int \frac{(h_\nu(x) - w_\nu) f(x;\nu + \theta) dx}{[h_\nu(x) \leq w_\nu]} \right) \]

\[= p + c_\nu + w_\nu + \min_{\nu} \left( (\theta + \nu)^2 + c\delta(\nu) + \int \frac{(h_\nu(x) - w_\nu) f(x;\nu + \theta) dx}{[h_\nu(x) \leq w_\nu]} \right) . \]

Since \( h_\nu(x) \) is symmetric about zero and increasing in \(|x|\), again by Appendix C, the integral is minimized for \( \nu = - \theta \). Therefore,

\[ R_{\nu+1}(\theta) = c_{\nu+1} + \min\{w_{\nu+1}, h_{\nu+1}(\theta)\} , \tag{4.17} \]

where \( c_{\nu+1} = p + c_\nu + w_\nu \),

\[ w_{\nu+1} = c + \int \frac{(h_\nu(x) - w_\nu) f(x;0) dx}{[h_\nu(x) \leq w_\nu]} \]

and

\[ h_{\nu+1}(\theta) = \theta^2 + \int \frac{(h_\nu(x) - w_\nu) f(x;\theta) dx}{[h_\nu(x) \leq w_\nu]} . \]

The constant \( w_{\nu+1} \) is associated with taking \( \nu = - \theta \), while \( h_{\nu+1}(\theta) \) is associated with the choice \( \nu = 0 \). Further,
\[ v_{\ell+1}(\theta) = \begin{cases} -\theta & \text{if } |\theta| \geq k_{\ell+1} \\ 0 & \text{if } |\theta| < k_{\ell+1} \end{cases} \]

where the critical constant \( k_{\ell+1} > 0 \) is determined by setting

\[ h_{\ell+1}(\theta) = w_{\ell+1} \]

or

\[ g_{\ell+1}(\theta) = h_{\ell+1}(\theta) - w_{\ell+1} = 0 \]

and solving for \( \theta \). Note that

\[ g_{\ell+1}(\theta) = h_{\ell+1}(\theta) - w_{\ell+1} \]

\[ = (\theta^2 - c) + \int_{[h_{\ell}(x) < w_{\ell}]} (h_{\ell}(x) - w_{\ell})[f(x;\theta) - f(x;0)]dx \]

\[ = (\theta^2 - c) + \int_{-k_{\ell}}^{k_{\ell}} g_{\ell}(x)[f(x;\theta) - f(x;0)]dx \]

where \( k_{\ell} \) satisfies \( g_{\ell}(k_{\ell}) = g_{\ell}(-k_{\ell}) = 0 \).

Since \( h_{\ell}(\theta) \) was assumed to be symmetric about zero and increasing in \( |\theta| \), \( h_{\ell+1}(\theta) \) (4.17) must be of the same form. By induction, the functional (4.15) must be of this form for every \( t \). Therefore, we have
Theorem 4.1. The optimal control policy for system (4.4) with performance criterion (4.13) is to take

\[ u_{n-\ell} = -\hat{\theta}_{n-\ell} \quad \text{if} \quad |\hat{\theta}_{n-\ell}| \geq k_{\ell} \quad (4.18) \]

\[ = 0 \quad |\hat{\theta}_{n-\ell}| < k_{\ell} \quad \ell = 1, 2, \ldots, n , \]

where \( k_{\ell} > 0 \) satisfies \( g_{\ell}(k_{\ell}) = g_{\ell}(-k_{\ell}) = 0 \), with

\[ g_{1}(\theta) = (\theta^2 - c) \quad (4.19) \]

\[ g_{\ell+1}(\theta) = (\theta^2 - c) + \int_{-k_{\ell}}^{k_{\ell}} g_{\ell}(x)(f(x;\theta) - f(x;0))dx \quad \text{for} \ \ell > 1 , \]

and \( \hat{\theta}_{t} \) is the generalized Kalman filter estimate of \( \theta_{t} \) given data through time \( t \).

Note that since \( g_{\ell}(x) \) is negative and symmetric about zero on \([-k_{\ell}, k_{\ell}]\), \( g_{\ell+1}(x) \) is also negative and symmetric about zero on \([-k_{\ell+1}, k_{\ell+1}]\). And since \( g_{\ell}(x) \) is increasing in \(|x|\), it follows that \( g_{\ell+1}(x) \) is increasing in \(|x|\). Also, from (4.19) and Appendix C, we can see that \( k_{\ell} \leq c_{1/2} \) for every \( \ell \). That is, the control action limits can never be greater than \( c_{1/2} \), where \( c \) is the cost ratio from (4.13).

The critical constants for the optimal control policy described by (4.18) and (4.19) can be approximated numerically, using a quadrature
approach. Note that the control action limits depend only on $c$, the cost ratio from (4.13), and $\sigma^2_y$, the system variance from (4.4). Table 4.1 gives the optimal control limits for special cases of the five period problem ($n=5$). Note that for the cases represented, the control limits appear to approach a limiting value.

We will later use a result from Bather (1963) to show that in general, the control limits converge to a limiting value as the number of control periods tends to infinity. Tables of these limiting values for various $(c, \sigma^2_y)$ combinations will also be presented.

Table 4.1. Critical constants for five period control problem, with system variance $\sigma^2_y = 1$

<table>
<thead>
<tr>
<th>$n$</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.500</td>
<td>.707</td>
<td>.866</td>
<td>1.000</td>
<td>1.414</td>
<td>1.732</td>
<td>2.000</td>
<td>2.236</td>
</tr>
<tr>
<td>2</td>
<td>.493</td>
<td>.683</td>
<td>.818</td>
<td>.925</td>
<td>1.223</td>
<td>1.428</td>
<td>1.595</td>
<td>1.740</td>
</tr>
<tr>
<td>3</td>
<td>.493</td>
<td>.683</td>
<td>.819</td>
<td>.929</td>
<td>1.230</td>
<td>1.432</td>
<td>1.587</td>
<td>1.715</td>
</tr>
<tr>
<td>4</td>
<td>.493</td>
<td>.683</td>
<td>.819</td>
<td>.929</td>
<td>1.230</td>
<td>1.432</td>
<td>1.589</td>
<td>1.717</td>
</tr>
<tr>
<td>5</td>
<td>.493</td>
<td>.683</td>
<td>.819</td>
<td>.929</td>
<td>1.230</td>
<td>1.432</td>
<td>1.589</td>
<td>1.718</td>
</tr>
</tbody>
</table>

E. Complete State Information

Consider again the linear system described by (4.4) with cost criterion (4.13). If $\sigma^2_e = 0$ in (4.4), the process or state mean is observed without error. In this case it is common to say we have
complete state information. With complete state information, (4.15) continues to hold, with \( p = \sigma_n^2 \). And the optimal \( u_t \)'s are defined by

\[
\begin{align*}
  u_{n-l} &= \nu_{l}(\theta_{n-l}) \\
  l &= 1, 2, \ldots, n.
\end{align*}
\]

The control action limits are the same as for cases where \( \sigma_n^2 \neq 0 \), but the control actions depend on the observed \( \theta_t \)'s rather than the filtered estimates \( \hat{\theta}_t \). The infinite horizon problem related to the complete state case is considered by Box and Jenkins (1963).

F. Convergence Results

Bather (1963) proves in a context more general than ours the uniform convergence of the sequence of functions \( R_n(\theta) - R_n(0) \) to a continuous bounded function, where \( R_n(\theta) \) is defined by (4.17). In the case being considered here,

\[
R_n(\theta) - R_n(0) = g_n(\theta) \cdot I[|\theta| \leq k_n] + c.
\]

The conditions for convergence given by Bather hold here, implying the uniform convergence of the sequence of functions \( g_n(\theta) \cdot I[|\theta| \leq k_n] \). This fact allows us to prove the convergence of the sequence of constants \( k_n \). That is, we have

**Proposition 4.1** Consider the sequence of functions \( g_n \) from (4.19). If the sequence \( g_n(\theta) \cdot I[|\theta| \leq k_n] \) converges uniformly, then
1) The $g_n$ sequence converges uniformly to a function $g$ and the sequence of critical constants $k_n$ converges to some value $k^*$ and

and 2) $g$ and $k^*$ are such that:

$$g(\theta) = (\theta^2 - c) + \int_{-k^*}^{k^*} g(x)K(\theta, x)dx$$

and $g(k^*) = 0$,

where $K(\theta, x) = f(x; \theta) - f(x; 0)$, for $f(x; \mu)$ the $N(\mu, \sigma^2)$ density.

Proof: Let $h(x)$ be the (uniform) limit of the sequence of functions $g_n(x)I[|x| \leq k_n]$ and let $g(\theta)$ be defined by

$$g(\theta) = (\theta^2 - c) + \int h(x)K(\theta, x)dx$$

Then clearly,

$$|g_n(\theta) - g(\theta)| \leq \int \left| (g_{n-1}(x)I[|x| \leq k_{n-1}] - h(x))K(\theta, x) \right| dx$$

$$\leq \sup_x \left| g_{n-1}(x)I[|x| \leq k_{n-1}] - h(x) \right| \int |K(\theta, x)| dx$$

$$\leq 2 \sup_x \left| g_{n-1}(x)I[|x| \leq k_{n-1}] - h(x) \right| ,$$

and $g_n$ converges uniformly to $g$. 

Next, notice that since each \( g_n \) is continuous and nondecreasing in \(|\theta|\), so also must be \( g \). Since \( g_n \to g \) uniformly, \( \exists N_\varepsilon \ni \) for every \( n \geq N_\varepsilon \),

\[
|g(k_n) - g_n(k_n)| < \varepsilon
\]
or

\[
|g(k_n) - 0| < \varepsilon
\]

That is, \( g(k_n) \to 0 \), so that

\[
g(\lim k_n) = g(\lim k_n) = 0 \text{ and in fact } g \text{ must be 0 on the}
\]
(possibly degenerate) interval \([\lim k_n, \lim k_n]\).

Now take \( k^* = \lim k_n \) and note that

\[
g_n(x)I[|x| \leq k_n] = (g_n(x) - g(x))I[|x| \leq k_n] + g(x)I[|x| \leq k_n]
\]

and that the first summand on the right converges uniformly to 0 and the second converges uniformly to \( g(x)I[|x| \leq k^*] \) so that

\[
h(x) = g(x)I[|x| \leq k^*]
\]

and we have

\[
g(\theta) = (\theta^2 - c) + \int_{-k^*}^{k^*} g(x)K(\theta, x)dx \quad . \quad (4.20)
\]
But, $\theta^2 - c$ is strictly increasing in $\theta$, and $\int_{-k^*}^{k^*} g(x)K(\theta,x)dx$ is nondecreasing in $\theta$ since $g(x)$ is nonpositive, symmetric about zero, and monotone in $|x|$ on $[-k^*, k^*]$. So, in fact $g(\theta) = (\theta^2 - c) + \int_{-k^*}^{k^*} g(x)K(\theta,x)dx$ is strictly increasing in $|\theta|$ on $[-k^*, k^*]$ and has a single positive root on that interval so that

$$k^* = \lim_{n \to \infty} k_n = \lim_{n \to \infty} k_n = \lim_{n \to \infty} k_n,$$

and the control action constants converge to a steady state value.

Numerical evidence (as in Table 4.1) suggests that the convergence need not be monotone.

G. Asymptotic Average Risk - Optimal Control Policy

For the finite horizon problem, minimization of the expected loss (4.13) over all control policies $\{u_t\}_{t=0}^{n-1}$ was the objective. This is clearly equivalent to minimization of $n^{-1}L(n)$. It is not only possible to determine the form of the optimal policy but also to identify the limit (as $n \to \infty$) of the optimal value of $n^{-1}L(n)$, the expected cost per stage using Bather's result and Proposition 4.1.

$R_n(\theta)$ (4.17) is the risk associated with the $n$ period optimal control policy described by (4.18, 4.19), given a start at $\hat{\theta}_0 = \theta$. 
From (4.17), applying the recursive relationships,

\[ R_n(\theta) = c_n + \min\{w_n, h_n(\theta)\} \]

\[ = p + c_{n-1} + w_{n-1} + w_n + (h_n(\theta) - w_n) \cdot I[h_n(\theta) \leq w_n] \]

\[ = c_{n-1} + w_{n-1} + (p+c) + \int_{-k_{n-1}}^{k_{n-1}} g_{n-1}(x)f(x;0)dx \]

\[ + g_n(\theta) \cdot I[|\theta| < k_n] \]

\[ \vdots \]

\[ = n(p+c) + \int_{-k_1}^{k_1} g_1(x)f(x;0)dx + \int_{-k_2}^{k_2} g_2(x)f(x;0)dx \]

\[ + \cdots + \int_{-k_{n-1}}^{k_{n-1}} g_{n-1}(x)f(x;0)dx + g_n(\theta) \cdot I[|\theta| < k_n] . \]

Using the uniform convergence of \( g_n(x) \cdot I[|x| \leq k_n] \) to \( g(x)I[|x| \leq k^*] \) established in Proposition 4.1, we have that

\[ \frac{1}{n} g_n(\theta) \cdot I[|\theta| \leq k_n] \to 0 \text{ uniformly on } (-\infty, +\infty), \]

and by the bounded convergence theorem
\[
\int_{-k}^{1-n} g_{n-1}(x)f(x;0)dx \to \int_{-k}^{k^*} g(x)f(x;0)dx.
\]

Using the fact that if a sequence converges so must its Cesaro mean, we then have that

\[
\lim_{n \to \infty} \frac{1}{n} R_n(\theta) = (p+c) + \int_{-k}^{k^*} g(x)f(x;0)dx, \quad (4.21)
\]

where \( g(x) = (x^2-c) + \int_{-k}^{k^*} g(y)K(x,y)dy \), and \( k^* = \lim_{n \to \infty} k_n \) is such that \( g(k^*) = 0 \).

H. Asymptotic Average Risk - Shewhart Control Policy

Both as a device for use in identifying the limiting value of the critical constants \( k_n \) defining an optimal control policy and because it can be of independent interest, define a "Shewhart" control policy by

\[
\begin{align*}
\hat{u}_t &= -\hat{\theta}_t \quad \text{if} \quad |\hat{\theta}_t| > k \\
&= 0 \quad \text{if} \quad |\hat{\theta}_t| < k,
\end{align*}
\]

119
where as usual $\hat{\theta}_t = E[\theta_t | y^t]$, and $k$ is a constant. This is similar to the approach taken by Box and Jenkins (1963) in their treatment of the complete state information problem.

Let

$$s_1(\theta) = \int (x^2 + c6(v))h(x;v+\theta)dx$$

where $h(x;u)$ is the $N(u,p)$ density, with $p$ defined by (4.6). Then

$$s_1(\theta) = p + (v+\theta)^2 + c6(v)$$

$$= p + \theta^2 \cdot I[|\theta| < k] + c \cdot I[|\theta| > k]$$

$$= (p+c) + (\theta^2-c) \cdot I[|\theta| < k] .$$

Note that under the Shewhart control policy,

$$L(1) = E\{\theta_1^2 + c6(u_0)\}$$

$$= p + (\hat{\theta}_0 + u_0)^2 + c6(u_0)$$

$$= (p+c) + (\hat{\theta}_0-c) \cdot I[|\hat{\theta}_0| < k]$$

$$= s_1(\hat{\theta}_0) .$$
Now, define recursively,

\[ s_\ell(0) = p + (\theta + v)^2 + c\delta(v) + \int_{-k}^{k} s_{\ell-1}(x)f(x;\nu+\theta)dx , \] (4.22)

where \( f(x;\nu) \) is the \( \mathcal{N}(\mu,\sigma^2) \) density, and note that under the Shewhart control policy

\[ L(n) = s_n(\hat{\theta}_0) . \]

This is the risk associated with the \( n \)-period Shewhart control policy described above, with performance criterion (4.13). We proceed to find an expression for \( s_n(\theta) \).

First, note that

\[ s_2(\theta) = p + (\theta + v)^2 + c\delta(v) + (p+c) + \int_{-k}^{k} (x^2 - c)f(x;\nu+\theta)dx \]

\[ = 2p + c + (c + \int_{-k}^{k} (x^2 - c)f(x;\nu)dx)\cdot I[|\theta| \geq k] \]

\[ + (\theta^2 + \int_{-k}^{k} (x^2 - c)f(x;\theta)dx)\cdot I[|\theta| < k] \]

\[ = 2(p+c) + \int_{-k}^{k} (x^2 - c)f(x;\nu)dx \]

\[ + ((\theta^2 - c) + \int_{-k}^{k} (x^2 - c)(f(x;\theta) - f(x;\nu))dx)\cdot I[|\theta| < k] . \]
Then, with \( g_1(x) = x^2 - c \) as before, define the linear operator \( \mathcal{L}_k \) by

\[
\mathcal{L}_k m(x) = \int_{-k}^{k} m(y)K(x,y)dy,
\]

where \( K(x,y) \) is defined by (4.20), and \( m(y) \) is a bounded continuous function on \([-k,k]\). Then, using this operator notation,

\[
s_2(\theta) = 2(p+c) + \int_{-k}^{k} g_1(x)f(x;0)dx + (g_1(\theta) + \mathcal{L}_k g_1(\theta)) \cdot I[|\theta| < k],
\]

and

\[
s_3(\theta) = 3(p+c) + 2 \int_{-k}^{k} g_1(x)f(x;0)dx + \int_{-k}^{k} \mathcal{L}_k g_1(x)f(x;0)dx
\]

\[
+ (g_1(\theta) + \mathcal{L}_k g_1(\theta) + \mathcal{L}_k^2 g_1(\theta)) \cdot I[|\theta| < k].
\]

Generalizing to the \( n \) period risk function of the Shewhart policy,

\[
s_n(\theta) = n(p+c) + \int_{-k}^{k} \sum_{i=0}^{n-2} (n-1-i) \mathcal{L}_k^i g_1(x)f(x;0)dx
\]

\[
+ (\sum_{i=0}^{n-1} \mathcal{L}_k^i g_1(\theta)) \cdot I[|\theta| < k].
\]

Consider the problem of determining the large \( n \) behavior of

\[
\frac{1}{n} s_n(\theta).
\] The following proposition will be used to prove the uniform
convergence of $\frac{1}{n} s_n(\theta)$ under the condition on $k$ and $\sigma_\|$ that with

$$I_m^m(\theta) \equiv \int_{-k}^{k} \cdots \int_{-k}^{k} |K(x_2, x_1)| dx_1 |K(x_3, x_2)| dx_2 \cdots |K(\theta, x_m)| dx_m,$$

$$\max_{|\theta| < k} I_m(\theta) < 1.$$ (4.24)

**Proposition 4.2** If condition (4.24) holds for some positive integer $m$, and $\mathcal{L}_k$ is the linear operator defined by (4.23), then,

1) The Neumann series $\sum_{i=0}^{n} \mathcal{L}_k^i g_1(x)$ converges uniformly on $[-k, k]$ to a function $g_k^*$ satisfying

$$g_k^*(x) = g_1(x) + \int_{-k}^{k} g_k^*(y) K(x, y) dy$$

and

2) $\sum_{i=0}^{n-2} \frac{(n-1-i)}{n} \mathcal{L}_k^i g_1(x)$ also converges uniformly to $g_k^*(x)$.

**Proof:**

Let

$$a = \max_{|x| < k} |g_1(x)|,$$

$$b = \max \left\{ \max_{|\theta| < k} I_1(\theta), \max_{|\theta| < k} I_2(\theta), \ldots, \max_{|\theta| < k} I_{m-1}(\theta) \right\}.$$
and \( r = \max_{|\theta| \leq k} \mathfrak{I}^m(\theta) < 1 \).

Then, for \( x \in [-k,k] \),

\[
|\mathcal{E}^n_k g_1(x)| \leq abr^\frac{n}{m}, \quad n = 1, 2, \ldots,
\]

where \([x]\) is the greatest integer function. The terms of the Neumann series

\[
g_1(x) + \mathcal{E}^n_k g_1(x) + \mathcal{E}^{2n}_k g_1(x) + \ldots + \mathcal{E}^{kn}_k g_1(x) + \ldots
\]

are thus uniformly bounded in absolute value by the terms of the convergent series of numbers

\[
a + abr^\frac{1}{m} + abr^\frac{2}{m} + \ldots + abr^\frac{n}{m} + \ldots,
\]

which assures the uniform absolute convergence of the Neumann series.

Now,

\[
\sum_{i=0}^{n} \mathcal{E}^i_k g_1(x) = g_1(x) + \mathcal{E}^{n-1}_k (\sum_{i=0}^{n} \mathcal{E}^i_k g_1)(x) \quad (4.25)
\]

and if \( g_k^* \) represents the limit of the Neumann series, the bounded convergence theorem shows that
So, taking limits in (4.25) we have that

\[ g_k^*(x) = g_1(x) + \mathcal{L}_k g_k^*(x) \]

for \( x \in [-k,k] \).

To prove (2), choose \( \lambda(e) \) large enough that

\[ \frac{ab(m)}{1-r} \left[ \frac{t-1}{m} \right] \leq \epsilon. \]

Then, for large \( n \),

\[
\begin{align*}
& \left| \sum_{i=0}^{n-2} \left( \frac{n-1-i}{n} \right) \mathcal{L}_k g_1(x) - g_k^*(x) \right| \\
& \leq \left| g_k(x) - \sum_{i=0}^t \mathcal{L}_k g_1(x) \right| \\
& \quad + \left| \sum_{i=0}^t \left( \frac{n-1-i}{n} \right) \mathcal{L}_k g_1(x) - \sum_{i=0}^t \mathcal{L}_k g_1(x) \right| \\
& \quad + \left| \sum_{i=t+1}^{n-2} \left( \frac{n-1-i}{n} \right) \mathcal{L}_k g_1(x) \right|.
\end{align*}
\]

Now, the first term satisfies

\[
\begin{align*}
& \left| g_k(x) - \sum_{i=0}^t \mathcal{L}_k g_1(x) \right| \\
& \leq ab \sum_{i=t+1}^{\infty} r^i
\end{align*}
\]

\[
\leq ab(m) \sum_{i=[\frac{t+1}{m}]}^{\infty} r^i
\]
The third term satisfies

\[
\left| \sum_{i=t+1}^{n-2} \frac{(n-1-i)}{n} \zeta_i g_1(x) \right| \leq ab \sum_{i=t+1}^{\infty} r^i \leq \varepsilon, \quad \text{for all } x \in [-k, k].
\]

Also,

\[
\left| \sum_{i=0}^{t} \frac{(-i-1)}{n} \zeta_i g_1(x) \right| \leq \frac{a(t+1)}{n} \max_{i \leq t} \left| \zeta_i g_1(x) \right| \quad \text{for all } x \in [-k, k],
\]

Thus, choosing \( n(t, \varepsilon) \) large enough that

\[
\frac{a(t+1)}{n} \max_{i \leq t} \left| \zeta_i g_1(x) \right| \leq \varepsilon
\]

completes the proof. \( \square \)

The standard operator notation for \( g_k^*(x) \) satisfying 1) is

\[
 g_k^*(x) = (I - \zeta_k)^{-1} g_1(x), \quad \text{where } I \text{ is the identity operator.}
\]

A numerical check shows that the condition \( \max_{|\theta| \leq k} I^1(\theta) < 1 \) holds.
for $k \leq 2.3\sigma_v$ and the condition $\max_{|\theta| \leq k} I^2(\theta) < 1$ holds for $k \leq 3.0\sigma_v$.

Using Proposition 4.2, as $n \to \infty$,

$$\frac{1}{n} s_n(\theta) \to (p+c) + \int_{-k}^k g_k^*(x)f(x;\theta)dx,$$  \hspace{1cm} (4.26)

where

$$g_k^*(x) = g_1(x) + \int_{-k}^k g_k^*(y)K(x,y)dy,$$

and the convergence is uniform on $(-\infty, +\infty)$. Clearly, to find the value of $k$ which minimizes the asymptotic average risk for Shewhart policies for which (4.26) holds, we must find a minimizer of

$$\int_{-k}^k g_k^*(x)f(x;\theta)dx.$$  \hspace{1cm} (4.27)

I. Evaluation of Steady State Control Action Limits

Comparing (4.21) and (4.26) it is clear that the formal expressions for the asymptotic average risk associated with the optimal control policy (4.17, 4.18) and the optimal asymptotic average risk for a Shewhart control policy are the same. The side conditions that, respectively, $k^* = \lim k_n$ satisfy $g(k^*) = 0$ and that an optimal $k$ minimize
\[ \int_{-k}^{k} g_k^*(x)f(x;0)\,dx, \quad \text{though, are somewhat different. However, at least} \]
in the case that \( k^* \leq 3.0\sigma_v \), since the hypotheses of Proposition 4.2
are satisfied for \( k \leq 3.0\sigma_v \), representation (4.26) will hold for
\( k = k^* \).

Using the \( m=2 \) version of condition (4.24) we thus know that
Shewhart policies with \( k \leq 3.0\sigma_v \) have convergent average risks given
by (4.26) and that amongst these policies, the one with \( k = k^* \) is
optimal. So, at least for cases where \( c_{l2} \leq 3.0\sigma_v \) (and so \( k^* \leq 3.0\sigma_v \)),
choice of \( k \) to minimize (4.27) on \([0,c_{l2}]\) will produce a value \( k_{\min} \) that
is both \( \lim k_n = k^* \) and best \( k \) for any Shewhart policy with convergent
average risk.

Proposition 4.1 can be used directly to numerically approximate
the value of \( k^* \in [0,c_{l2}] \) for which \( (3-L_+)^{-1} g_1(k^*) = 0 \). Or numerical
methods can be used to approximate the value of \( k \) that minimizes (4.27),
giving (at least when \( c_{l2} \leq 3.0\sigma_v \)) the optimal value of the Shewhart
control limit and the limit of the \( k_n \)'s for the overall optimal policy.
Numerical evidence not given in this thesis strongly suggests that for
\( c \) as large as \( 10,000\sigma_v^2 \) the minimizer of the formal expression (4.27)
over the interval \([0,c_{l2}]\) is also \( k^* = \lim k_n \).

Note that if we let \( k_0(c,\sigma_v^2) \) represent the steady state control
limit for a system with cost ratio \( c \) and system variance \( \sigma_v^2 \), a simple
change of variables in (4.21) shows that
\[ k_0(c, \sigma_{\nu}^2) = \sigma_{\nu} \cdot k_0(c/\sigma_{\nu}^2, 1). \]

So, only steady state values associated with \( \sigma_{\nu}^2 = 1 \) need be tabled. Table (4.2) gives an extensive list of steady state control limits, determined using the results of Proposition 4.1. These values agree to two decimal places with a limited number of values given by Box and Jenkins (1963), obtained in an entirely different way.

<table>
<thead>
<tr>
<th>( c )</th>
<th>( k_0(c, 1) )</th>
<th>( c )</th>
<th>( k_0(c, 1) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.9283</td>
<td>600</td>
<td>7.1487</td>
</tr>
<tr>
<td>2</td>
<td>1.2294</td>
<td>700</td>
<td>7.4534</td>
</tr>
<tr>
<td>3</td>
<td>1.4321</td>
<td>800</td>
<td>7.7271</td>
</tr>
<tr>
<td>4</td>
<td>1.5884</td>
<td>900</td>
<td>7.9762</td>
</tr>
<tr>
<td>5</td>
<td>1.7175</td>
<td>1000</td>
<td>8.2054</td>
</tr>
<tr>
<td>10</td>
<td>2.1657</td>
<td>2000</td>
<td>9.8726</td>
</tr>
<tr>
<td>20</td>
<td>2.6966</td>
<td>3000</td>
<td>10.9902</td>
</tr>
<tr>
<td>30</td>
<td>3.0521</td>
<td>4000</td>
<td>11.8545</td>
</tr>
<tr>
<td>40</td>
<td>3.3268</td>
<td>5000</td>
<td>12.5691</td>
</tr>
<tr>
<td>50</td>
<td>3.5537</td>
<td>6000</td>
<td>13.1833</td>
</tr>
<tr>
<td>60</td>
<td>3.7488</td>
<td>7000</td>
<td>13.7248</td>
</tr>
<tr>
<td>70</td>
<td>3.9208</td>
<td>8000</td>
<td>14.2109</td>
</tr>
<tr>
<td>80</td>
<td>4.0751</td>
<td>9000</td>
<td>14.6535</td>
</tr>
<tr>
<td>90</td>
<td>4.2156</td>
<td>10000</td>
<td>15.0605</td>
</tr>
<tr>
<td>100</td>
<td>4.3447</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>5.2841</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>5.9137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>6.4005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>6.8029</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
V. BIBLIOGRAPHY


Roberts, S. W. 1959. Control chart tests based on geometric moving averages. Technometrics, 1, 239-250.


VI. ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Dr. Stephen B. Vardeman for his guidance in this research and for his assistance in the writing of this thesis.

Thanks to Darlene Wicks for the careful and patient typing of this thesis.

Finally, I wish to thank my wife, Lee Ann, for her continued patience and encouragement during the writing of this thesis.
Lemma 2.1 Let

\[ q_t = (q_{t-1} + b)(\frac{a}{a+b+q_{t-1}}) \]

where \( a > 0 \) and \( b > 0 \).

Then, \( \lim_{t \to \infty} q_t = (-b + (b^2 + 4ab)^{1/2})/2 \).

Proof: Note that

\[ |q_t - q_{t-1}| = |(q_{t-1} + b)(\frac{a}{a+b+q_{t-1}}) - (q_{t-2} + b)(\frac{a}{a+b+q_{t-2}})|. \]

Suppose first that \( q_{t-1} > q_{t-2} \). Then, the first term is larger than the second, and

\[ |q_t - q_{t-1}| \leq |q_{t-1} + b)(\frac{a}{a+b+q_{t-2}}) - (q_{t-2} + b)(\frac{a}{a+b+q_{t-2}})| \]

\[ \leq |q_{t-1} - q_{t-2}| \cdot \frac{a}{a+b+q_{t-2}} \]

\[ \leq |q_{t-1} - q_{t-2}| \cdot \frac{a}{a+b} \]
where

\( \left( \frac{a}{a+b} \right) < 1 \).

Suppose \( q_{t-1} < q_{t-2} \). Then, the first term is smaller than the second, and both are positive, so again

\[
|q_t - q_{t-1}| \leq |(q_{t-1} + b)(\frac{a}{a+b+q_{t-2}}) - (q_{t-2} + b)(\frac{a}{a+b+q_{t-2}})|
\]

\[
\leq |q_{t-1} - q_{t-2}| \left( \frac{a}{a+b+q_{t-2}} \right)
\]

\[
\leq |q_{t-1} - q_{t-2}| \cdot \left( \frac{a}{a+b} \right).
\]

So, we have

\[
|q_t - q_{t-1}| \leq |q_{t-1} - q_{t-2}| \left( \frac{a}{a+b} \right) \text{ for } \left( \frac{a}{a+b} \right) < 1,
\]

and the convergence of the \( q_t \) sequence follows. Let \( \lim_{t \to \infty} q_t = q \). Then, \( q \) must satisfy

\[
q = (q+b) \left( \frac{a}{a+b+q} \right).
\]

So,

\[
q(a+b+q) = (q+b)a,
\]

\[
q^2 + bq - ab = 0,
\]
and $$q = \frac{-b \pm (b^2 + 4ab)^{1/2}}{2}.$$ Since $$q$$ must be positive, we take the positive root, and

$$q = \frac{-b + (b^2 + 4ab)^{1/2}}{2}. \quad \square$$

Notice that with $$a = \sigma_c^2$$ and $$b = \sigma_y^2$$, $$q_t = V(\theta_t | y^t)$$ for the model specified by (4.4). So, the lemma can be applied to show the convergence of the variance of the posterior distribution of the current $$\theta_t$$ given the current data.

B. Convergence of $$\alpha_t$$ Sequence

Lemma 2.2 Suppose that a sequence $$\{\alpha_t, t = 1, 2, \ldots\}$$ satisfies

$$\alpha_t = a + b \alpha_{t-1},$$

for $$0 < b < 1$$, $$a > 0$$. Then, $$\lim_{t \to \infty} \alpha_t = a/(1-b).$$

Proof:

$$|\alpha_t - \alpha_{t-1}| = |b(\alpha_{t-1} - \alpha_{t-2})| \leq |\alpha_{t-1} - \alpha_{t-2}| \cdot b$$

Since $$|b| < 1$$, the sequence must converge. If $$\alpha = \lim_{t \to \infty} \alpha_t$$, then $$\alpha$$ must satisfy
\[ \alpha = a + b\alpha \]
or \[ \alpha = a/(1-b) . \]

C. A Monotonicity Result

Lemma 4.1 Suppose \( X \sim N(\theta, \sigma^2) \) and \( h(x) \) is symmetric about zero, decreasing in \( |x| \), and positive for \( |x| < c_0 \), where \( c_0 \) is a positive constant. Then, \( E_{\theta}\{h(X) \cdot I[|X| \leq c_0]\} \) is nonincreasing in \( |\theta| \).

Proof: Let \( Y = h(X) \cdot I[|X| \leq c_0] \). Then, \( Y \geq 0 \), and

\[
E_{\theta}(h(X) \cdot I[|X| \leq c_0]) = E_{\theta}(Y) \\
= \int_0^\infty P[Y > t]dt \\
= h(0) \int_0^\infty P_\theta[h(X) > t]dt ,
\]

since the maximum of \( h(x) \) is \( h(0) \).

For any \( t_0 \in [0, h(0)] \),

\[
P_\theta[h(X) \geq t_0]
\]
= P_\theta [-k_0 \leq X \leq k_0], \text{ where } h(k_0) = t_0.

Since \( X \sim N(\theta, \sigma^2) \), this probability is monotonically nonincreasing in
|\theta|. Now, pick \( 0 \leq \theta_1 \leq \theta_2 \). Then,

\[
E_{\theta_1} (h(X) \cdot I[|X| < c_0]) = \int_0^{h(0)} P_{\theta_1}[h(X) \geq t] dt
\]

\[
\geq \int_0^{h(0)} P_{\theta_2}[h(X) \geq t] dt
\]

= \( E_{\theta_2} (h(X) \cdot I[|X| < c_0]) \),

since \( P_{\theta_1}[h(X) \geq t] \geq P_{\theta_2}[h(X) \geq t] \forall t \in [0, h(0)] \).