Title: Empirical Studies on the Estimation of Variance in Linear Regression Models

Abstract approved: Donald Guthrie, Jr.

An iterative approach is suggested for the estimation of the error covariance matrix $\Sigma$ to find approximate BLUE estimators in linear regression models. It is shown through experimental studies how the variances can be estimated in the simple general linear regression model, in the linear regression model sequenced over time, and in the linear regression model with a stochastically varying parameter vector, commonly called the Kalman model. All studies assume the true error covariance matrix $\Sigma$ to be diagonal.

The general linear regression model is shown to be a special case of the Kalman model. Bounds are found for the covariance matrix of the parameter vector estimation error of the Kalman model. They show the parameter vector estimator in the Kalman model to be an inconsistent estimator of the true parameter vector with consistency applying only when the true parameter vector is nonrandom.
EMPIRICAL STUDIES ON THE ESTIMATION
OF VARIANCES IN LINEAR REGRESSION MODELS
by
JAMES ELROY REINMUTH

A THESIS
submitted to
OREGON STATE UNIVERSITY

in partial fulfillment of
the requirements for the
degree of
DOCTOR OF PHILOSOPHY
June 1970
APPROVED:

Redacted for Privacy

Associate Professor of Mathematics and Statistics
in charge of major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented       June 12, 1969

Typed by Charlene Laski for     James Elroy Reinmuth
ACKNOWLEDGEMENT

The author extends his thanks and appreciation to Dr. Donald Guthrie, Jr. for his help and advice in the preparation of this dissertation. Dr. Guthrie has been an inspiration to the author throughout his graduate program and his advice and guidance is largely responsible for the academic proficiency the author now enjoys.

Also, the author had the benefit of the valuable advice of Dr. W. Scott Overton and Dr. Don Pierce of Oregon State University and Dr. David B. Duncan of The Johns Hopkins University throughout the preparation of the study herein. This project could not have been undertaken without the financial support provided by the National Science Foundation. To all the aforementioned professors, The National Science Foundation, and unnamed professors and graduate students of the Department of Statistics at Oregon State University the author will be forever grateful for the advice, criticism, and support they have given him in his graduate studies.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 The general linear model</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Properties of the weighted least squares estimators</td>
<td>6</td>
</tr>
<tr>
<td>1.2 The iterative least squares estimation technique</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Estimation of the error covariance matrix</td>
<td>12</td>
</tr>
<tr>
<td>1.3.1 The case $\sum = \theta \beta \gamma$, $\beta \gamma$ known</td>
<td>12</td>
</tr>
<tr>
<td>1.3.2 The case where $\sum$ is arbitrary</td>
<td>13</td>
</tr>
<tr>
<td>1.3.3 The case of stationary, correlated errors</td>
<td>17</td>
</tr>
<tr>
<td>1.3.4 The case of uncorrelated errors, an approximate approach</td>
<td>19</td>
</tr>
<tr>
<td>1.3.5 Variance estimation in practice</td>
<td>21</td>
</tr>
</tbody>
</table>

| 2.0 Recursive models | 23 |
| 2.1 Incorporation of prior information | 24 |
| 2.1.1 The partitioned general linear model | 25 |
| 2.1.2 The Bayesian formulation | 28 |
| 2.2 The linear model as a stochastic process | 31 |
| 2.2.1 The Kalman model | 33 |
| 2.2.2 The relation between the Kalman model and the nonstochastic linear model | 36 |
| 2.2.3 Asymptotic properties of the stochastic model | 38 |
| 2.3 Covariance matrix estimation in the recursive model | 51 |
| 2.3.1 Covariance matrix estimation when the parameter vector is nonrandom or constant | 51 |
| 2.3.2 Covariance matrix estimation in the Kalman model | 53 |
### Table of Contents (con't)

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0 Experimental studies</td>
<td>56</td>
</tr>
<tr>
<td>3.1 Estimation of the covariance matrix</td>
<td>57</td>
</tr>
<tr>
<td>(diagonal) from one observation vector</td>
<td></td>
</tr>
<tr>
<td>3.1.1 Choice of the smoothing constant</td>
<td>58</td>
</tr>
<tr>
<td>3.1.2 Distribution of the iterative variance estimators</td>
<td>61</td>
</tr>
<tr>
<td>3.1.3 Goodness of the estimators</td>
<td>65</td>
</tr>
<tr>
<td>3.2 Variance estimation in the recursive model</td>
<td>67</td>
</tr>
<tr>
<td>3.2.1 The diagonal approximation method</td>
<td>68</td>
</tr>
<tr>
<td>3.2.2 The general smoothing method</td>
<td>70</td>
</tr>
<tr>
<td>3.2.3 Comparison of weighted least squares estimators</td>
<td>73</td>
</tr>
<tr>
<td>3.2.4 Discussion of the experimental results</td>
<td>98</td>
</tr>
<tr>
<td>3.3 Variance estimation in the Kalman model</td>
<td>101</td>
</tr>
<tr>
<td>3.3.1 Variance estimation</td>
<td>101</td>
</tr>
<tr>
<td>3.3.2 Detection of the Kalman model</td>
<td>103</td>
</tr>
<tr>
<td>3.3.3 Dependence on initial values</td>
<td>107</td>
</tr>
<tr>
<td>3.4 Conclusion</td>
<td>112</td>
</tr>
</tbody>
</table>

Appendix 1

Sufficiency of the prior estimator for the current state in the recursive models.

Appendix 2

The Bayesian formulation for the recursive models.

Appendix 3

Behavior of the parameter vector estimator in the recursive models.

Appendix 4

Duncan's theorem equating the Kalman model and the general linear model.

Bibliography 134
1.0 The General Linear Model

In the general linear model, we assume a dependent variable $Y$ is representable as the sum of a linear function of $p$ predictor or independent variables, and a random disturbance. This random disturbance will be called the observational error, or error of the regression equation. For a sample of $n$ such dependent variables, we may write the model as

$$Y = X\beta + \varepsilon$$

where $Y$: (nx1), $X$: (nxp), $\beta$: (px1), $\varepsilon$: (nx1) and the matrix $X$ is of rank $p < n$. The design matrix $X$ and the parameter vector $\beta$ will be regarded as non-random. The error vector is assumed to satisfy the properties

$$E(\varepsilon) = 0, \quad E(\varepsilon \varepsilon') = \Sigma.$$ 

We shall later add the additional assumption of normality of the observational errors, but shall retain generality in our early discussion.

Under the general linear model, we find the best, linear, unbiased estimator (BLUE) of the regression para-

---

1. The notation $M$: (rcx) is used to indicate that $M$ is a matrix with $r$-rows and $c$-columns.
meter vector \( \hat{\beta} \) as the vector which minimizes the quadratic function

\[
(1.2) \quad Q(\hat{\beta}) = (Y - X\hat{\beta})' \sum^{-1}(Y - X\hat{\beta}).
\]

This BLUE weighted least squares estimator is

\[
\hat{\beta}(\sum) = (X'\sum^{-1}X)^{-1}X'\sum^{-1}Y.
\]

For a discussion, see Deutsch (6), page 61.

Weighted least squares estimators were first discussed by Aitken (1) who assumed the error covariance matrix \( \sum \) to be known. Since complete knowledge of the matrix \( \sum \) is seldom available in practice, we must be willing to accept an estimator for the parameter vector whose covariance matrix is less positive definite than that for \( \hat{\beta}(\sum) \).

The most commonly used estimator for the parameter vector is the linear least squares estimator

\[
(1.3) \quad \hat{\beta}(LS) = (X'X)^{-1}X'Y
\]

which minimizes the quadratic function

\[
(1.4) \quad Q(\hat{\beta}) = (Y - X\hat{\beta})'(Y - X\hat{\beta}).
\]

The geometrical justification for the use of estimator (1.3) is well known. Suppose we let \( X \) be a matrix whose column vectors are \( x_1, x_2, \ldots, x_p \). Then we could write

\[
x\hat{\beta} = \beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p
\]
as a point in the linear subspace of Euclidean n-space spanned by \( x_1, x_2, \ldots, x_p \). The vectors \( x_1, x_2, \ldots, x_p \) are called the regression vectors and the linear subspace is called the regression space \( M(X) \). Dropping a perpendicular from the point in n-space with position \( \overline{Y} \) onto \( M(X) \) gives \( \hat{X}(LS) \) as the vector at the foot of that perpendicular. That is, if
\[
\hat{Y} = \hat{X}
\]
for an arbitrary estimator \( \hat{\beta} \), then the distance\(^2\)
\[
||\overline{Y} - \hat{\overline{Y}}||^2
\]
has a minimum which is attained when and only when \( \hat{X}(LS) \) is the projection of \( \overline{Y} \) on \( M(X) \), where \( \hat{\beta}(LS) \) is the least squares estimate of \( \beta \). This is clearly shown in diagram 1 on the following page.

This perpendicular is the residual vector \( e \), lying in the orthogonal complement of \( M(X) \), which satisfies
\[
(1.5) \quad \overline{Y} = \hat{X}(LS) + e.
\]

In the true model (1.1), the error vector is
\[
e = \overline{Y} - \hat{\beta}.
\]

2. The length of a vector \( \overline{Y} \) is denoted by \( ||\overline{Y}|| \).
Thus, we could write

\[ \varepsilon = (X\hat{\beta}(LS) - X\beta) + \varepsilon \]

or

\[ \varepsilon = \varepsilon^* + \varepsilon \]

which shows that the residual errors \( \varepsilon \) from the least squares estimate \( \hat{\beta}(LS) \) are affected by \( \varepsilon^* \), where \( \varepsilon^* \) belongs to the regression space \( M(X) \).

Diagram 1

![Diagram](image)

Figure 1.1 illustrates that the distance \( \| Y - \hat{Y} \|^2 \) between the observation vector \( Y \) and \( \hat{Y} = X\hat{\beta} \), for arbitrary \( \hat{\beta} \), is a minimum if and only if \( \hat{Y} \) is the orthogonal projection of \( Y \) on \( M(X) \). This projection is unique but, \( \hat{\beta}(LS) \) is unique if and only if \( X \) is of full rank.
The preceding discussion also applies to the weighted least squares problem. Suppose we let $P$ be a nonsingular $(n \times n)$ matrix such that $P^T P' = I$. If we let $Z = P Y$, then

$$E(Z) = PE(Y) = PX \hat{\beta}.$$  

The quadratic function $Q(\hat{\beta})$ from (1.2) is

$$Q(\hat{\beta}) = (Y - X \hat{\beta})' \Sigma^{-1} (Y - X \hat{\beta})$$

$$= (Z - PX \hat{\beta})' (Z - PX \hat{\beta}).$$

To this last expression we could apply the geometric arguments of the previous discussion which show that the expression is minimized by

$$\hat{\beta}(\Sigma) = ((PX)'(PX))^{-1}(PX)'Z$$

$$= (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Y.$$

The least squares estimators $\hat{\beta}(LS)$ are optimum in a minimum variance sense (BLUE) if the errors are stationary and uncorrelated and $\Sigma$ is a constant multiple of the identity matrix. The relationships among the columns of the design matrix $X$ and the eigenvectors of the error covariance matrix $\Sigma$ and the conditions on the eigenvalues of $\Sigma$, required for $\hat{\beta}(LS) = \hat{\beta}(\Sigma)$, are discussed in a recent paper by Watson (19).
For nonstationary and/or correlated errors, the estimation of $\beta$ by some weight matrix $\omega$ and then the estimation of $R$ by weighted least squares is often preferable, in the sense of providing an estimator with a smaller covariance matrix, to the estimation of $\beta$ by $\hat{\beta}(LS)$. This non BLUE weighted least squares estimator, where we let the weight matrix $\omega$ replace the unknown covariance matrix $\Sigma$, is

$$\hat{\omega}(\omega) = (X'\omega^{-1}X)^{-1}X'\omega^{-1}y.$$  

One can see that $\hat{\omega}(\Sigma)$ and $\hat{\omega}(LS)$ are both special cases of $\hat{\omega}(\omega)$.

To find a good weight matrix $\omega$ of $\Sigma$, one must first estimate $\beta$ from an assumed weight matrix and then use the corrected residuals to find a new weight matrix. Hence, we are suggesting that an iterative technique finding first an estimator for $\beta$ and then for $\Sigma$ will, after a sequence of steps, provide improvement over the least squares estimators $\hat{\beta}(LS)$.

1.1 Properties of the Weighted Least Squares Estimators

Under the assumptions of the general linear model, the weighted least squares estimator $\hat{\beta}(\omega)$ is unbiased for any choice of weight matrix $\omega$. This can be shown if we write
\[ \hat{\beta}(\omega) = (X'\omega^{-1}X)^{-1}X'\omega^{-1}Y \]

\[ = (X'\omega^{-1}X)^{-1}X'\omega^{-1}(X\hat{\beta} + \varepsilon) \]

\[ = (X'\omega^{-1}X)^{-1}X'\omega^{-1}X\hat{\beta} + (X'\omega^{-1}X)^{-1}X'\omega^{-1}\varepsilon \]

\[ = \hat{\beta} + (X'\omega^{-1}X)^{-1}X'\omega^{-1}\varepsilon. \]

If the matrix \( \omega \) is fixed, then since \( X \) is nonrandom and \( E(\varepsilon) = 0 \),

\[ E(\hat{\beta}(\omega)) = E(\hat{\beta}) = \beta. \]

It must be noted that if we employ an iterative estimation scheme where the weight matrix \( \omega \) is based on the residuals from a previous iterative step, then the estimator \( \hat{\beta}(\omega) \) is not unbiased. In this case \( \omega \) is not fixed but is a function of the products of the observations. When the weight matrix \( \omega \) and the error vector \( \varepsilon \) are dependent, the estimator \( \hat{\beta}(\omega) \) is still unbiased provided that \( E(\varepsilon|\omega) = 0 \). Even though the estimator \( \hat{\beta}(\omega) \) may be biased, the effect of the bias is an iterative scheme will be negligible. The estimation problem is then one of examining how much better is \( \hat{\beta}(\hat{\omega}) \) than any other weighted least squares estimator \( \hat{\beta}(\omega) \).
Suppose we let the covariance matrix of $\hat{\beta}(\omega)$ be represented by $\Psi(\omega)$. Then, we could write

$$
\Psi(\omega) = E \left\{ (\hat{\beta}(\omega) - \beta)(\hat{\beta}(\omega) - \beta)' \right\} = E \left\{ (QY - \beta)(QY - \beta)' \right\}
$$

where $Q = (X'\omega^{-1}X)^{-1}X'\omega^{-1}$. Our expression reduces to

$$
\Psi(\omega) = QE \left\{ (X\beta + \varepsilon)(X\beta + \varepsilon)' \right\} Q'
$$

$$
= QX\beta\beta' - \beta\beta'X'Q + \beta\beta'
$$

$$
= QX\beta\beta'X'Q + QE(\varepsilon\varepsilon')Q'
$$

$$
- QX\beta\beta' - \beta\beta'X'Q + \beta\beta'
$$

$$
= QE(\varepsilon\varepsilon')Q'
$$

$$
= Q \sum Q'
$$

(1.1.1)

$$
= (X'\omega^{-1}X)^{-1} X'\omega^{-1} \sum \omega^{-1}X(X'\omega^{-1}X)^{-1}.
$$

In the case where $\omega = \sum$, this expression becomes

(1.1.2)

$$
\Psi(\sum) = (X'\sum^{-1}X)^{-1}.
$$

Grenander and Rosenblatt (10), using an inequality for matrices related to the Schwartz inequality for functions, have shown that for any weight matrix $\omega$, ...
That is, the covariance structure \( \psi(\omega) \) is minimized by the choice of weight matrix \( \omega = \frac{1}{T} \).

When we use the linear least squares estimators, expression (1.1.1) becomes

\[
\psi(\text{LS}) = (X'X)^{-1}X'\Psi X(X'X)^{-1}.
\]

The least squares error covariance matrix may or may not be smaller than \( \psi(\omega) \). The general procedure for comparing the efficiency of \( \hat{\beta} \) (LS) and \( \hat{\beta} \) (\( \omega \)) to the best estimator \( \hat{\beta} \) (\( \Psi \)) is offered by expression (1.1.3), and by Watson (19).

1.2 The Iterative Least Squares Estimation Technique

As noted above, estimation of the parameter vector \( \beta \) is necessary as a prerequisite for the estimation of \( \Psi \) regardless of its form. This may be accomplished in an iterative manner where we begin with a workable estimate of \( \beta \), perhaps \( \beta \) (LS). Then we use the corrected residuals to estimate \( \Psi \). The estimate of the covariance matrix \( \Psi \) obtained is then used as the weight matrix for finding the parameter vector estimator at the second step. This procedure is repeated for a predetermined number of

---

3. A positive definite matrix \( M \) is less than or equal to a positive definite matrix \( N \) if the difference \( D = N - M \) is positive semi-definite.
steps or until the difference between successive estimates is less than a given value.

Mallios (13) gives a sufficient condition for the convergence of a sequence of regression parameters in an iterative scheme where the weight matrix employed is arbitrary. In a later section discussing the results of experimental studies, closeness of the iterative estimates for $\hat{\beta}$ and $\hat{\beta}(\beta_{l})$ to $\hat{\beta}(\beta)$ and $\hat{\beta}$, respectively, is measured by the size of the Euclidean norm of their differences.

In a sequential outline form, the iterative least squares estimation technique progresses as follows:

**Step 1.** We obtain the linear least squares estimator

$$\hat{\beta}(\text{LS}) = (X'X)^{-1}X'Y.$$  

Let $\hat{\beta}(\text{LS}) = \hat{\beta}_{1}$.

**Step 2.** Using a corrected form of the residuals $Y - X\hat{\beta}_{1}$, calculate an estimator for the error covariance matrix. Let this estimator be

$$\hat{\Sigma}_{1} = \sigma_{1}.$$

**Step 3.** Using $\omega_{1}$ as the weight matrix, calculate

$$\hat{\beta}_{2} = (X'(\omega_{1})^{-1}X)^{-1}X'(\omega_{1})^{-1}Y.$$  

**Step 4.** In an iterative manner, return to Step 2 to find $\hat{\Sigma}_{2} = \omega_{2}$. Then proceed to Step 3 and

---

4. The Euclidean norm of an (rxc) matrix $M = \{m_{ij}\}$ is

$$||M||_{E} = \left( \sum_{i=1}^{r} \sum_{j=1}^{c} m_{ij}^2 \right)^{1/2}.$$
repeat the process for a predetermined number of iterative cycles.

At the end of \( k \)-cycles of the iterative process, our weighted least squares estimator of \( \beta \) is

\[
\hat{\beta}_k = (X'(\omega_{k-1})^{-1}X)^{-1}X'(\omega_{k-1})^{-1}Y, \text{ for } k = 2, 3, \ldots
\]

An estimator of the covariance matrix of the iterative estimator's estimation error is given by

\[
\hat{\Psi}(\hat{\beta}_k) = (X'(\omega_{k-1})^{-1}X)^{-1}X'(\omega_{k-1})^{-1}\omega_k(\omega_{k-1})^{-1}X(X'(\omega_{k-1})^{-1}X)^{-1}.
\]

The principle of the iterative weighted least squares method is to provide good estimators of \( \hat{\beta} \) and then \( \hat{\beta} \), based upon all available information regarding \( \hat{\beta} \) and \( \hat{\beta} \). We do not seek estimators of the regression parameter vector \( \beta \) which are good in the sense of being closest the quantity \( \beta \) being estimated. What we seek is an estimator whose covariance matrix is smaller than that of any other linear estimator we might consider. That is, through the iterative weighted least squares technique, we obtain an accurate estimate of the covariance matrix \( \hat{\Psi} \). Then, by employing the weighted least squares estimator (1.2.1), we are returned an estimator for \( \hat{\beta} \) which is close to the best, linear, unbiased estimator \( \hat{\beta}(\hat{\Psi}) \).

5. Within the experimental studies of chapter 3, the closeness of two matrices is measured by the magnitude of the Euclidean norm of their difference.
1.3 Estimation of the Error Covariance Matrix

The selection of the proper estimation scheme to use in step 2 of the iterative approach is dependent upon the characteristics of the design matrix $X$ and the error vector $\xi$ as well as certain statistical considerations. In the process of estimating the parameter vector $\beta$ by $\hat{\beta}(\omega)$, we are imposing $p$ restrictions on the $n$ observations. Thus, if the error covariance matrix $\sum$ is to be estimated from a function of the estimators $\hat{\beta}(\omega)$, we have at most $n-p$ degrees of freedom available. Since $\sum$ has in its most general form $n(n+1)/2$ elements, it may then be very difficult to find an estimator for $\sum$ unless we are willing to make certain structural assumptions. This section will examine variance estimation techniques which might be considered for step 2 of the iterative process when something of the form of $\sum$ is assumed known.

1.3.1 The Case $\sum = \theta M$, $M$ Known

In this case the covariance matrix is entirely known except for the scalar factor $\theta$. The matrix $M$ is assumed to be symmetric, positive semi-definite, and nonsingular. The $n$-dimensional linear model with independent observations whose errors have a common variance is a special case of this section.

Suppose we seek the BLUE of the parameter vector $\beta$. 
From expression 1.2, we know that

\[ \hat{\theta}(\theta M) = (X'(\theta M)^{-1}X)^{-1}X'(\theta M)^{-1}Y \]

\[ = \theta (X'M^{-1}X)^{-1} \frac{1}{\theta} X'M^{-1}Y \]

\[ = (X'M^{-1}X)^{-1}X'M^{-1}Y \]

since \( \theta \) is a scalar. Therefore, \( \hat{\theta}(\theta M) = \hat{\theta}(M) \).

If we let

\[ \tilde{Y} = M^{-1/2}Y \quad \text{and} \quad \tilde{X} = M^{-1/2}X \]

then it follows that an unbiased estimator for the scalar parameter \( \theta \) is provided by the least squares theory as

\[ \hat{\theta} = (\tilde{Y} - \tilde{X}\hat{\theta}(M))' (\tilde{Y} - \tilde{X}\hat{\theta}(M)) / n-p, \]

where the estimator \( \hat{\theta}(M) \) does not involve the unknown factor \( \theta \). A proof of this may be found in either Cramer (5) or Scheffe (16).

1.3.2 The Case Where \( \gamma \) Is Arbitrary

As we mentioned earlier, when \( p \) regression parameters are estimated from \( n \) observations, the \( n \) residuals calculated from the observations are associated with only \( n-p \) degrees of freedom. Thus, THE COMPUTED RESIDUALS ARE NOT INDEPENDENT EVEN THOUGH THE TRUE ERRORS, AND HENCE THE OBSERVATIONS, ARE INDEPENDENT.
If the model is

\[ Y = X\beta + \varepsilon \]

\[ E(\varepsilon) = 0, \quad E(\varepsilon \varepsilon') = \sum, \quad X'X \text{ nonsingular} \]

then the residuals can be written as

\[ e = Y - \hat{Y} = Y - X\hat{\beta}(\omega) \]

\[ = Y - X(X'\omega^{-1}X)^{-1}X'\omega^{-1}Y \]

\[ = (I - X(X'\omega^{-1}X)^{-1}X'\omega^{-1})Y \]

\[ = (I - K)Y , \]

where the matrix \( K = X(X'\omega^{-1}X)^{-1}X'\omega^{-1} \). Since \( Y = X\beta + \varepsilon \), it follows that

\[ e = (I - K)(X\beta + \varepsilon) \]

\[ = X\beta + \varepsilon - X(X'\omega^{-1}X)^{-1}X'\omega^{-1}(X\beta + \varepsilon) \]

\[ = X\beta + \varepsilon - X\beta - Ke \]

(1.3.2) \( = (I - K)e \)

which shows the relationship between the residuals and the true error structure of the model. From (1.3.2), we can see that the second order sample moments of the residuals
are
\[ e' e' = (I - K) e' e' (I - K)' \]

which has mathematical expectation

\[(1.3.3) \quad E(e' e') = (I - K) \sum (I - K)'. \]

It has been noted earlier that when we employ an iterative scheme the weight matrix \( \omega_{\lambda} \) and, hence, the matrix

\[ K_{\lambda} = X(X' \omega_{\lambda}^{-1} X)^{-1} X' \omega_{\lambda}^{-1} \]

are functions of previous corrected residuals and are technically not constant. Suppose we disregard this fact, whose resultant bias is negligible in the iterative scheme anyway, and consider \( \omega_{\lambda} \) and \( K_{\lambda} \) as constant for the \( \lambda \)th iterative step. Then if we let

\[ V(e_{\lambda}) = (Y - X \hat{\omega}_{\lambda})(Y - X \hat{\omega}_{\lambda})' \]

where \( \hat{\omega}_{\lambda} \) is defined in section 1.2, from (1.3.3)

\[ E(V(e_{\lambda})) = (I - K_{\lambda}) \sum (I - K_{\lambda})'. \]

Furthermore, suppose that we assume the weight matrix at the \( \lambda \)th iteration satisfies

\[(1.3.4) \quad \omega_{\lambda} = \sum + F_{\lambda} \]

and that the residual cross products satisfy

\[(1.3.5) \quad V(e_{\lambda}) = (I - K_{\lambda}) \sum (I - K_{\lambda})' + G_{\lambda} \]
where the elements of $F_\ell$ and $G_\ell$ have expectation zero.

Exponential smoothing gives an estimate $\omega_{\ell+1}$ of
based on both the past estimate $\omega_\ell$ and the new residual
cross product information as

$$\omega_{\ell+1} = \omega_\ell + \Lambda \{ V(\epsilon_k) \} + (I - \Lambda) \omega_\ell$$

$$= \omega_\ell + \Lambda \{ V(\epsilon_k) - \omega_\ell \}$$

where the eigenvalues of $\Lambda$ are less than 1 in absolute
value to guarantee convergence. The estimate $\omega_{\ell+1}$ is a
weighted average of $\omega_\ell$ and $V(\epsilon_k)$.

In order to obtain an unbiased estimator for the true
covariance matrix $\mathbf{\Sigma}$, we would revise our estimator
$\omega_{\ell+1}$ to read

$$(1.3.6) \quad \omega_{\ell+1} = \omega_\ell + \Lambda \{ V(\epsilon_k) - (I - K_\ell) \omega_\ell (I - K_\ell)' \}.$$  

The best smoothing matrix $\Lambda$ involves a function of the
variances of $F_\ell$ and $G_\ell$. In the absence of the know-
ledge of these variances, we would use as an estimate of $\Lambda$

$$\hat{\Lambda} = \theta I,$$

where $0 \leq \theta \leq 1$ to guarantee convergence. The scalar
term $\theta$ can be allowed to vary from one iterative step to
another (decreasing from 1 to 0) to reflect the degree

6. The best smoothing matrix is the one which would mini-
mize the covariance matrix of the difference $(\omega_{\ell+1} - \mathbf{\Sigma})$. 

of confidence, measured by $1 - \theta$, in the previous estimator $\omega_k$. Chapter 3 discusses the rate of convergence of $\omega_k$ to $\omega_0$ for different choices of the scalar $\theta$. It will be noted later that the unbiased estimator (1.3.6) is the same as the one we would obtain by minimizing the quadratic risk function of $F_\ell$ and $G_\ell$ with respect to $\omega_0$. This topic is extensively discussed in Chapter 2, section 1.

1.3.3 **The Case of Stationary, Correlated Errors**

For the general linear model discussed in section 1.0 and referenced as (1.1) suppose the $n$-observations of the observation vector $Y$ are taken at different and equally spaced times. Furthermore, suppose the observations and hence, the errors $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are correlated and stationary over time. Then the elements of the error covariance matrix $\Sigma$ are a function only of the separation time and

$$
\Sigma =
\begin{bmatrix}
\sigma(0) & \sigma(1) & \cdots & \sigma(n-1) \\
\sigma(1) & \sigma(0) & \cdots & \sigma(n-2) \\
\vdots & \vdots & \ddots & \vdots \\
\sigma(n-1) & \sigma(n-2) & \cdots & \sigma(0)
\end{bmatrix}
$$

That is, the $i,j$th element of $\Sigma$, which is $\sigma_{ij} = \sigma(t)$, is a function of only $t = |i-j|$. 

An estimator of \( \sigma(t) \) to use at the \( \ell \)th step of the iterative scheme is

\[
(1.3.7) \quad \hat{\sigma}(t) = \frac{1}{n-t} \sum_{i=1}^{n-t} (y_i - \hat{Y}_{\ell-1,i})(y_{i+t} - \hat{Y}_{\ell-1,i+t}).
\]

for \( t = 0, 1, \ldots, n-1 \), and \( \ell = 1, 2, \ldots \) and where

\[
y_i = \text{the } i\text{th element of } \bar{Y}, \quad \text{and} \quad \hat{Y}_{\ell-1,i} = \text{the } i\text{th element of } \hat{Y}_{\ell-1} = X\hat{\theta}_{\ell-1}.
\]

The covariance matrix at the \( \ell \)th iterative step is then

\[
(1.3.8) \quad \hat{C}_\ell = \begin{bmatrix}
\hat{\sigma}(0) & \hat{\sigma}(1) & \ldots & \hat{\sigma}(n-1) \\
\hat{\sigma}(1) & \hat{\sigma}(0) & \ldots & \hat{\sigma}(n-2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}(n-1) & \hat{\sigma}(n-2) & \ldots & \hat{\sigma}(0)
\end{bmatrix}
\]

where the elements of \( \hat{C}_\ell \) are defined by (1.3.7).

Estimator (1.3.6) may be adapted to further correct the residuals at each iteration. Under the assumptions that

\[
E(\omega_\ell) = \underline{I}, \quad \text{and} \quad E(\hat{C}_\ell) = \underline{I}
\]

for all \( \ell \), the smoothed estimator

\[
(1.3.9) \quad \omega_{\ell+1} = \Lambda (\hat{C}_\ell) + (I - \Lambda) \omega_\ell
\]
is an unbiased estimator for \( \hat{\Sigma} \) provided that the eigenvalues of \( \Lambda \) are less than 1 in absolute value. Either estimator (1.3.8) or (1.3.9) is superior to (1.3.6) for the estimation of the covariance matrix \( \hat{\Sigma} \) in this case since (1.3.6) does not guarantee that the elements along each diagonal of \( \omega_k \) will be equal. Estimators (1.3.8) and (1.3.9) incorporate this additional assumption that the errors are stationary.

1.3.4 The Case of Uncorrelated Errors, An Approximate Approach

Suppose the covariance matrix \( \hat{\Sigma} \) is assumed to be diagonal with the diagonal elements not necessarily equal. If the observations are taken over time, the errors will be assumed nonstationary.

Suppose we have used the optimum weights in obtaining the estimator for \( \hat{\Sigma} \). Then, the residuals are

\[
e = \bar{Y} - \hat{\Sigma} = \bar{Y} - X\hat{\Sigma}(\hat{\Sigma}^{-1}) \\
= (I - X(X'X)^{-1}X')(\hat{\Sigma}^{-1} \text{Y}) \\
= (I - K)\text{Y},
\]

where the matrix \( K = X(X'X)^{-1}X' \hat{\Sigma}^{-1} \). The residual cross products have mathematical expectation

\[
E(e e') = (I - K)\hat{\Sigma}(I - K)' \\
= \hat{\Sigma} - 2K \hat{\Sigma} + K\hat{\Sigma}K \\
= \hat{\Sigma} - 2K \hat{\Sigma} + K\hat{\Sigma} \]
since \( K'K = K \), when \( K \) is defined as above.

Thus,

\[
(1.3.10) \quad E(e_i e'_j) = (I - K) \Sigma
\]

when the weight matrix \( w = \Sigma \).

If we ignore the residual cross product terms \( e_i e_j \) where \( i \neq j \), expression (1.3.10) leads us to the approximate residual equations

\[
e_i^2 = (1 - K_{ii}) \sigma_i^2, \quad i = 1, 2, \ldots, n
\]

where \( K_{ii} \) is the \( i, i \)th element of the matrix \( K \), and \( \sigma_i^2 \) is the \( i \)th element of the diagonal matrix \( \Sigma \). This suggests the choice of

\[
(1.3.11) \quad \hat{\sigma}_i^2 = \frac{e_i^2}{(1 - K_{ii})}, \quad i = 1, 2, \ldots, n
\]

as an estimator for the \( i \)th diagonal element of \( \Sigma \). The covariance matrix is estimated by

\[
(1.3.12) \quad \hat{\Sigma} = \begin{bmatrix}
\hat{\sigma}_1^2 & 0 & \ldots & 0 \\
0 & \hat{\sigma}_2^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \hat{\sigma}_n^2
\end{bmatrix}
\]

In practice, the optimal weights are not known and (1.3.11) provides only approximately unbiased estimators
for the elements of $\hat{\Sigma}$. The advantage of this estimation technique is that it is computationally simple to apply. However, by employing an iterative scheme, either using (1.3.12) directly as the weights at each step or smoothing them and using (1.3.9) to estimate the weights, the bias due to these approximations can be reduced to negligible amounts.

1.3.5 Variance Estimation In Practice

The obvious shortcoming of all the suggested covariance estimation techniques is that each is based on only one set of data. As awkward as it may be, this is often the case in practice. The data analyst may have one set of data from an experiment which for economic or practical reasons cannot be repeated. Thus, he must make the most with what he has and, using all the information, both prior and experimental available to him, estimate the unknown characteristics from this one set of data.

The general relation between the expectations of the residual cross products and the true covariance matrix $\Sigma$ has been shown to be

$$E(e e') = (I - K) \hat{\Sigma} (I - K)'$$

If the matrix $K$ were of full rank, the solution for $\hat{\Sigma}$ would be unique. However, it was shown that $K$ is of rank $n-p$ so $(I - K)$ and $(I - K)'$ are singular matrices.
In practice, the analyst often times has intuitive knowledge about his experimental process which may be included as additional, independent restrictions. If enough of this prior, intuitive information is available, the matrix $K$ may be transformed to become nonsingular and give us the unique unbiased estimator

$$\hat{\Sigma} = (I - K)^{-1}(e^e')(I - K')^{-1}$$

for the error covariance matrix $\Sigma$.

The problem of solving for the unknown variance-covariance components at one time point is not strictly a mathematical problem. When we see the analyst as both the one who designs the experiment and interprets the data, it is apparent that he might know something about the process which is not readily quantifiable. This nonquantitative prior information should be incorporated into the analysis in the sense of adding additional restrictions to the data. The resultant conclusions then reflect all the analyst's available information, both quantitative and intuitive.

The case when a repeated experiment may be performed by the analyst is examined in the next chapter.
2.0 Recursive Models

Often, one discovers that the estimated values obtained from a set of experimental data vary more than some expected amount from the assumed correct values. This may be the result of a disproportionately large sampling variance or an indication of a process with a variable state structure. Not knowing which cause is responsible for this difference, the analyst will most likely repeat the original measurements hoping to obtain more accurate results.

Ordinary regression procedures applied to this new data provide only an independent estimator of the unknown process values while not incorporating the influence of the previous data. Thus, the data analyst may wish to perform a modified experiment to provide a check on the original data set as well as to provide an estimator based on all available data, both new and old. Such an experiment combines the original and new data sets into one model for estimating the unknown process parameters.

If such a process is repeated, we are suggesting a recursive estimation model. With this procedure, we start by making an initial estimate of the system parameters either from a minimal data set or by a judicious guess. Then each new data point is combined with the current set by an appropriate weighting scheme giving an updated
estimate of the system parameters. As the recursive scheme progresses, damping out the influence of remote estimators, it may be terminated at any convenient step offering an estimator based on all accumulated data up to that time. Furthermore, under assumptions of normality, the recursive state system estimator is sufficient for all the past data in estimating the current state at any point in the sequential process.\textsuperscript{1} This offers opportunity for significant economy in data storage.

2.1 Incorporation of Prior Information

Suppose our sample information for the new data set in the recursive process is represented by the model

\begin{equation}
Y = X\beta + \epsilon
\end{equation}

where the error structure satisfies the properties

\[ E(\epsilon) = 0, \quad E(\epsilon \epsilon') = \Sigma \]

the matrix \( X \) is of rank \( p \) and

\( Y: (nx1), \quad \beta: (px1). \)

In addition to the new sample information, we have prior information about the process derived from earlier samples in the recursive estimation scheme. Suppose our

\textsuperscript{1} See appendix 1.
previous sample data gave us the estimator $\hat{\beta}^*$ for the unknown state parameter vector $\hat{\beta}$. Then, our prior model is

\begin{equation}
(2.1.2) \quad \hat{\beta}^* = \hat{\beta} + \eta
\end{equation}

where $\eta$ is the estimation error from all past data which satisfies the properties

$$E(\eta) = 0, \quad E(\eta \eta') = I, \quad \text{and} \quad E(\eta \varepsilon') = 0.$$ 

In the recursive models, we often call the parameter vector the system's state vector. THE STATE OF THE SYSTEM IS THE STATE OF NATURE OF THE SYSTEM PARAMETERS AT SOME CYCLE OF THE RECURSION PROCESS. If our recursion involves data sequenced over evenly spaced time intervals and the error vectors $\varepsilon$ and $\eta$ follow a normal distribution, we have the Kalman recursive model (12). This model, to be discussed later, is especially important since it allows the state parameters to vary randomly over time.

2.1.1 The Partitioned General Linear Model

If we wish to take into account both the new and old information in our estimation technique, we combine (2.1.1) and (2.1.2) and write

\begin{equation}
(2.1.3) \quad \begin{bmatrix} Y \\ \hat{\beta}^* \end{bmatrix} = \begin{bmatrix} X \\ \mathbf{I} \end{bmatrix} \hat{\beta} + \begin{bmatrix} \varepsilon \\ \eta \end{bmatrix}.
\end{equation}
Since \( n \) and \( \varepsilon \) are independent due to their being selected from different experiments, the covariance structure for the combined model (2.1.3) is

\[
E \left( \begin{bmatrix} \varepsilon \\ n \end{bmatrix} \right) = \begin{bmatrix} \sum & 0 \\ 0 & \Gamma \end{bmatrix}.
\]

The combined model (2.1.3) is a partitioning of the general linear model. By applying generalized least squares to (2.1.3) we find that the BLUE for \( \beta \) based on both the sample and prior information is

\[
\hat{\beta}(\sum, \Gamma) = (X' \sum^{-1} X + \Gamma^{-1})^{-1} (\Gamma^{-1} \beta^* + X' \sum^{-1} Y)
\]

whose covariance matrix is

\[
\psi(\hat{\beta}(\sum, \Gamma)) = E[(\hat{\beta}(\sum, \Gamma) - \beta)(\hat{\beta}(\sum, \Gamma) - \beta')]
\]

\[
= (X' \sum^{-1} X + \Gamma^{-1})^{-1}.
\]

Suppose we rewrite the parameter vector estimator as

\[
\hat{\beta}(\sum, \Gamma) = \beta^* + \psi(\Gamma^{-1} \beta^* + X' \sum^{-1} Y) - \beta^*
\]

where, from (2.1.5), \( \psi = \psi(\hat{\beta}(\sum, \Gamma)) = (X' \sum^{-1} X + \Gamma^{-1})^{-1} \). Further simplifying, we have
\(\hat{\beta}(\bar{X}, \Gamma) = \beta^* + \psi X' \bar{X}^{-1}Y - \psi (\psi^{-1} - \Gamma^{-1} - 1) \beta^*\)

\(= \beta^* + \psi X' \bar{X}^{-1}Y - \psi (X' \bar{X}^{-1} + 1) \beta^*\)

(2.1.6) \(= \beta^* + \psi X' \bar{X}^{-1} (Y - X\beta^*)\).

Alternative form (2.1.6) is suggestive of an exponential smoothing type model with smoothing matrix \((\psi X' \bar{X}^{-1})\). Jones (11) has investigated this approach as applied to multivariate time series data. We have seen the exponential smoothing model as a variance estimation technique.

It is now apparent how our recursive process progresses. Each time we incorporate a new set of data into our estimation scheme, our state system estimator \(\hat{\beta}(\bar{X}, \Gamma)\) is

\[\hat{\beta}(\bar{X}, \Gamma) = \beta^* + \xi,\]

(2.1.7) where

\[\xi = \psi X' \bar{X}^{-1} (Y - X\beta^*)\]

\[= \psi X' \bar{X}^{-1} (\epsilon - \bar{X} \hat{\eta})\]

has expectation zero.² That is, we do not expect our new data to significantly alter our estimator for the parameter vector from the value calculated from all past data. This

². See appendix 3.
is true as long as our model is (2.1.3) containing

\[ \hat{\beta}^* = \hat{\beta} + \eta \]

as our prior information and

\[ Y = X\hat{\beta} + \epsilon \]

as the representation of our new sample data where the error vectors \( \eta \) and \( \epsilon \) satisfy earlier discussed properties. Equation (2.1.7) applies whether the state system parameters are random, nonrandom, or constant.

One can see from equations (2.1.6) and (2.1.7) that the recursive estimation procedure damps out the effect of remote data. Thus, if we begin with poor estimates of the process parameters, the recursive estimates should still converge toward the true process parameters after updating with additional experimental data.

The convergence properties of recursive estimators will be examined in detail in a later section.

2.1.2 The Bayesian Formulation

Further insight into the theory of recursive estimation is provided by a Bayesian estimation representation. If we make the following assumptions:

i) The regression parameter \( \hat{\beta} \) is a random vector

3. See appendix 2.
ii) Random vectors \( \varepsilon \) and \( \eta \) are normally distributed.

iii) \( \beta^* \) represents our prior estimate and \( Y \) represents our new experimental information.

Then the mean of the posterior distribution \( f(\beta|Y) \) is the least squares estimator (2.1.4). The mean of \( f(\beta|Y) \) is the Bayes estimate for a squared error loss function. The details of this discussion are carried out in appendix 2.

Furthermore, we note that if we assume the prior information is uniform in the sense of the prior covariance matrix \( \Gamma \) becoming very large, then the estimator is

\[
(2.1.8) \quad E(\beta|Y) = (X' \sum^{-1} X)^{-1} X' \sum^{-1} Y.
\]

That is, due to the lack of precision in the prior information, the prior estimate \( \beta^* \) is totally discounted in our estimator (2.1.8).

Intuitively speaking, this implies that if our measurements are extremely accurate, the error covariance matrix \( \sum \) will be small in some sense. Suppose we say

\[
\sum = \delta \sum_0
\]

where \( \delta \) is a small positive scalar quantity. Then, from (2.1.5)

\[
\psi = (\Gamma^{-1} + \frac{1}{\delta} X' \sum_0^{-1} X)^{-1}
\]
\[
\delta \left( \delta \delta^{-1} + x' \Sigma_0^{-1} x \right)^{-1} \\
= \delta \left( x' \Sigma_0^{-1} x \right)^{-1}.
\]

Therefore, when our measurements are very accurate, the state vector estimator

\[
\hat{\beta} = \delta \left( x' \Sigma_0^{-1} x \right)^{-1} \left( \Gamma^{-1} \beta^* + \frac{1}{\delta} x' \Sigma_0^{-1} y \right)
\]

\[
= \delta \left( x' \Sigma_0^{-1} x \right)^{-1} \frac{1}{\delta} \left( \delta \Gamma^{-1} \beta^* + x' \Sigma_0^{-1} y \right)
\]

\[
= (x' \Sigma_0^{-1} x)^{-1} (x' \Sigma_0^{-1} y)
\]

which has little dependence on the prior information.

On the other hand, if our prior intuition tells us that our measurements are very inaccurate, then \( \Sigma \) is quite large in some sense. We might say

\[
= \frac{1}{\delta} \Sigma_0^{-1}
\]

for \( \delta \) small but positive. Then,

\[
\Sigma^{-1} = \delta \Sigma_0^{-1}
\]

which is very small. This implies that our state vector estimator

\[
\hat{\beta} = (\Gamma^{-1})^{-1} (\Gamma^{-1} \beta^*)
\]

\[
= \beta^*
\]
which completely discounts the new experimental data.

This discussion of the Bayes concept can be applied to the estimation of the smoothing constant in the iterative schemes of Chapter 1. For instance, if at the $\lambda$ th iterative cycle we believe the weight matrix $\omega_\lambda$ to be very accurate, we let

$$\Lambda = \theta I$$

where $\theta$ is a small positive scalar. Then at the next iterative step the contribution of the new residual cross product information would be discounted more than if $\Lambda$ were held constant. If $\omega_\lambda$ is thought to be inaccurate, the choice of

$$\Lambda = \frac{1}{\theta} I$$

would provide the best weight.

2.2 The Linear Model As a Stochastic Process

Frequent applications of the linear regression model involve its extension over time. Signal detection and separation in control theory and economic prediction and forecasting offer common examples of a time dependent model. In both areas of application the stochastic linear regression model poses problems which do not exist in a nonstochastic experimental problem.

Implicit in most estimation techniques is the assump-
tion that the measurement errors are stationary, that the distribution function of the errors is independent of time. We say the errors are wide-sense stationary if their covariances and correlations depend only on the separation in time of selection between their corresponding observations.

Some physical processes cannot safely be assumed to possess stationary errors. In the tracking of missiles by radar systems, the received signals vary as the fourth power of the distance between the radar system and missile. Since the range and angle measurements are a function of these received signals, the covariance matrix of the measurement errors will vary along the missile's trajectory.

In some stochastic models the state parameters are random functions of time. This extension of the linear regression model provides a general framework for all linear estimation processes but is difficult to apply since the random excitations of the state parameters are unestimable. As an example of linear stochastic process with random state parameters, consider the following economic equation dealing with demand for textiles in the economy of a certain locality. The equation has the form

$$\log C_t = \beta_0 + \beta_1 \log P_t + \beta_2 \log M_t + \epsilon_t$$

where $C_t$ is the per capita textile consumption in year $t$,
P is the deflated price index of textiles, and M is the real per capita income. The regression parameters represent demand elasticities with \( \beta_1 \) representing the price elasticity and \( \beta_2 \) the income elasticity. Under varying governmental economic policies, these elasticity factors are not constant and should be subscripted as a function of time.

Kalman (12) gives an explicit formulation for a linear stochastic model with a completely random state system. Although his paper provides a generalized form for linear estimation problems it is restricted in its usefulness since it assumes the covariance matrices for the observational errors and state system excitations are known. In practice, these matrices are unknown.

2.2.1 The Kalman Model

Suppose that our process is stochastic over evenly spaced time points \( t = \Delta, 2\Delta, \ldots, T\Delta \). At any time \( t \), this linear stochastic process can be represented by two recursive equations. The observation equation expresses our new experimental data while the state equation represents the distributional pattern of the parameter vector over time.

4. The elasticity of demand for an item is the per cent change in the quantity demanded divided by the per cent change in price that the change in quantity demanded brings about.
The mathematical model for the Kalman system is

\[ y_t = X_t \hat{\beta}_t + u_t \]  \hspace{1cm} (2.2.1)

\[ \hat{\beta}_t = T_t \hat{\beta}_{t-1} + v_t \]

where \( Y_t: (nx1), \quad \beta_t: (px1) \) and the other terms are of corresponding dimensions. The Kalman model is a multivariate Markov process which may be stationary or nonstationary. The error vector \( u_t \) is assumed \( N(0, R_t) \) while the state system excitation \( v_t \) is \( N(0, Q_t) \). These random disturbance vectors are assumed to be independent at different time points as well as independent of each other.

Schweppe (17) has pointed out that general linear regression is a special case of the Kalman recursive theory. It is also apparent how this theory is useful for updating the estimates of the regression coefficients as more data becomes available.

For the Kalman model, the predicted state based on all the past data

\[ \hat{\beta}_t \big| t-1 = T_t \hat{\beta}_{t-1} \]

is sufficient from the past data for estimating the current

---

5. The terminology \( z \) is \( N(\mu, \nu) \) implies that \( z \) is a normally distributed vector with mean vector \( \mu \) and variance-covariance matrix \( \nu \).
state $\beta_t$. Furthermore, the estimation error $(\hat{\beta}_t | t-1 - \beta_t)$ is $N(0, S_{t|t-1})$ where

$$S_{t|t-1} = E(\hat{\beta}_t | t-1 - \beta_t)(\hat{\beta}_t | t-1 - \beta_t)'$$

$$= T_t S_{t-1} T_t' + Q_t$$

and where $S_{t-1} = E(\hat{\beta}_t-\beta_t)(\hat{\beta}_t-\beta_t)'$. The Kalman model combines the past estimate of the state vector $\hat{\beta}_t | t-1$ with the new experimental data $Y_t$ to find the BLU of the state vector $\beta_t$ based on all available data.

The resulting estimator, which is the Bayesian estimator for the state at time-\(t\), is

$$\hat{\beta}_t = [S_{t|t-1}^{-1} + X_t R_t^{-1} X_t']^{-1} [S_{t|t-1}^{-1} \hat{\beta}_t | t-1 + X_t R_t^{-1} Y_t].$$

Since $S_t = E(\hat{\beta}_t - \beta_t)(\hat{\beta}_t - \beta_t)'$

$$= [S_{t|t-1}^{-1} + X_t R_t^{-1} X_t]^{-1}$$

we can write (2.2.3) in the more convenient form

$$\hat{\beta}_t = S_t [S_{t|t-1}^{-1} \hat{\beta}_t | t-1 + X_t R_t^{-1} Y_t]$$

$$= \hat{\beta}_t | t-1 + S_t X_t R_t^{-1} (Y_t - X_t \hat{\beta}_t | t-1).$$

With knowledge of the error covariance matricies $R_t$ and $Q_t$, the estimator $\hat{\beta}_t$ is the best linear estimator of $\beta_t$ based

6. See appendix 1.
on all available data. However, \( \hat{\beta}_t \) is not unbiased since the true state is a random vector.\(^7\) The best we can say is that \[ E(\hat{\beta}_t) = E(\beta_t). \]

2.2.2 The Relation Between the Kalman Model and The Nonstochastic Linear Model

The essential difference between the partitioned general linear (PGL) model and the Kalman model is that the underlying state vector is assumed to be a normally distributed random vector in the Kalman model but is assumed nonrandom or constant in the ordinary general linear model. The PGL model and the Kalman model are identical if we make the equivalences:

\[
\begin{array}{c|c}
\text{Kalman} & \text{PGL} \\
\hline
\hat{\beta}_t & \hat{\beta}_t \\
\hat{\beta}_t | t-1 & \hat{\beta}_t | t-1 \\
R_t & R_t \\
S_t | t-1 & S_t | t-1 \\
Q_t & Q_t \\
S_t & S_t \\
T_t & T_t \\
T_t \hat{\beta}_{t-1} + v_t & v_t \\
u_t & e_t \\
v_t & 0 \\
\end{array}
\]

7. See appendix 3.
The Kalman model is very useful in the situation where a variable state structure may be present. It essentially gives the state vector elements a prior distribution at each time point based on past or subjective information. If we are willing to make two modifications to the Kalman model, a theorem by Duncan.(8) shows that we can view the Kalman model as a partitioned general linear model and apply ordinary weighted least squares theory for the estimation of the state vector $\hat{\theta}_t$. The two modifications required are that:

i) The prior means of the state vector $\hat{\theta}_t$ be considered as additional observations, and

ii) The covariance matrix of the estimates be that of the estimation error $\hat{\theta}_t - \theta_t$ and not of $\hat{\theta}_t$ alone.

The theorem states that the observation equation

$$y_t = X_t \hat{\theta}_t + u_t$$

and the prior information regarding the state

$$\hat{\theta}_t \mid t-1 = T_t \hat{\theta}_{t-1} = \hat{\theta}_t + \xi_t$$

where $u_t$ is $N(0, R_t)$ and $\xi_t$ is $N(0, S_t \mid t-1)$ may be written as a partitioned model in the form

8. For an explicit formulation and proof of Duncan's theorem, see appendix 4.
Then we can consider the state vector $\hat{\beta}_t$ as nonrandom in the usual sense of the general linear model and the weighted least squares estimator for $\hat{\beta}_t$ from (2.2.6) is the Kalman estimator (2.2.3).

Duncan's theorem is useful since it shows us that even if the state is considered random, the model may be reformulated such that in the revised model the state is nonrandom. Ordinary weighted least squares may then be employed to estimate the state from this revised model. Except for the time dimensions, our reformulated model (2.2.6) is identical to the partitioned general linear model with constant state vector referenced as (2.3).

2.2.3 Asymptotic Properties of the Stochastic Model

For the purpose of discussion, let us assume that our recursive model has a nonrandom state vector. Then our recursive equations are

\[
\begin{align*}
\text{OBS. EQ.} & \quad Y_t &= X_t\beta_t + u_t \\
(2.2.7) & \quad \text{STATE EQ.} & \quad \beta_t &= T_{t-1}\beta_{t-1}.
\end{align*}
\]

Furthermore, we shall assume the errors $u_t$ are independent, that $u_t$ is $N(0, R_t)$, $X_t$ is of full rank, $T_t$ is
nonsingular, and that the initial state vector estimate \( \hat{\beta}_0 \) is selected randomly with covariance matrix \( S_0 \).

Since \( \beta_t = T_{t-1} \beta_{t-1} \) expresses the state vector in a first order autoregressive form, we can write

\[
\hat{\beta}_t = T_t \beta_{t-1} \\
= T_t (T_{t-1} \beta_{t-2}) \\
= \phi_t, \beta_{t-2}
\]

where \( \phi_{i,j} = T_{i} \cdot T_{i-1} \cdots T_{j+1} , \phi_{i,i} = I \). Therefore, the observation vector can be written in terms of any state vector as

\[
Y_t = X_t \phi_{t,l} \beta_{l} + u_t \\
= H_t, \beta_{l} + u_t
\]

where \( H_{t,l} = X_t \phi_{t,l} \).

The least squares estimator for the state vector \( \beta_t \) is the familiar

\[
\hat{\beta}_t = \hat{\beta}_{t|t-1} + S_t X_t R_t^{-1} (Y_t - X_t \hat{\beta}_{t|t-1})
\]

However, the covariance matrix \( S_t \) is of a different form from (2.2.3) in the nonrandom case due to the absence of the

---

9. Since the only difference between the completely random model and our model with a nonrandom state vector is that in the latter \( Q_t = 0 \), our state vector estimator \( \beta_t \) is the same as (2.2.3).
state vector error covariance matrix $Q_t$.

At the first time period, the covariance matrix for the estimation error is

$$S_1 = [S^{-1}_1 0 + X_1^{-1}R_1^{-1}X_1]^{-1}$$

$$= [(T_1S_0T_1')^{-1} + X_1^{-1}R_1^{-1}X_1]^{-1}.$$

With the data from the next time period, the covariance matrix becomes

$$S_2 = [T_2(T_1S_0T_1')^{-1} + X_1^{-1}R_1^{-1}X_1]^{-1}T_2 + X_2^{-1}R_2^{-1}X_2]^{-1}$$

$$= [T_2(T_1S_0T_1')^{-1}T_2 + T_2X_1^{-1}R_1^{-1}X_1T_2 + X_2^{-1}R_2^{-1}X_2]^{-1}.$$  

By an inductive process,

$$(2.2.8) \quad S_t = [\Phi_{t,2}(T_1S_0T_1')^{-1}\Phi_t + \sum_{i=1}^{t} H_i'^{-1}R_i^{-1}H_i]^{-1}.$$

Suppose we let

$$K_t = \sum_{i=1}^{t} H_i'^{-1}R_i^{-1}H_i.$$  

This matrix is positive definite and satisfies the difference equation

$$K_{t+1} = \Phi_{t,t+1}^{-1}K_t\Phi_{t,t+1} + X_{t+1}^{-1}R_{t+1}^{-1}X_{t+1}.$$
Suppose we define the matrix $K_t^*$ as

$$K_t^* = K_1$$

$$K_t^* = \phi_{2,t} K_{t-2,t}; \text{ for } t \geq 2.1$$

Then, expression (2.2.8) may be written as

$$S_t = \left[\phi_{t,2} (T_1 S_0 T_1)^{-1} \phi_{t,2} + \phi_{t,2} K^* \phi_{t,2}\right]^{-1}$$

$$= \left[\phi_{t,2} (T_1 S_0 T_1)^{-1} + K^* \phi_{t,2}\right]^{-1}.$$

The matrix $K^*_t$ satisfies the difference equation

$$K^*_{t+1} = K^*_t + \phi_{2,t+1} X_{t+1} R_{t+1}^{-1} X_{t+1}^\prime \phi_{2,t+1}$$

A sufficient condition for the inverse matrix $(K^*_t)^{-1}$ to converge to the null matrix is that $K^*_{t+1} > K^*_t$ and that for some positive integer $q$, the partial sum of

$$\phi_{2,1} X_{i} R_{i}^{-1} X_{i} \phi_{2,1}$$

over any $q$ consecutive time intervals is positive definite.\(^{11}\) Since $R_i$ for all $i$ is an assumed positive

10. $\phi$ is the state transition matrix which equates

$$\beta_t = \phi_{t,\ell} \beta_{\ell}$$

for $\ell \leq t$. If we assume $\phi$ is nonsingular,

$$\beta_{\ell} = (\phi_{t,\ell})^{-1} \beta_t = \phi_{\ell,t} \beta_t.$$
definite covariance matrix, since \( \Phi_{2,i} \) is nonsingular, and since \( X_i \) is of full rank, then the partial sum

\[
\Phi_{2,i} X_i R_{11}^{-1} X_i \Phi_{2,i}
\]

over any \( q \) consecutive time intervals is positive definite. Thus,

\[
K_{t+j}^* > K_t^* \quad \text{for any} \quad j \geq q
\]

and the inverse \((K_t^*)^{-1}\) converges to the null matrix as the recursions \( t \) increase.

Applying the Cauchy-Schwartz inequality to the Euclidean norm of expression (2.2.9), we have

\[
\| S_t \|_E \leq \| \Phi_{2,t} \|_E^2 \| (T_1 S_0 T_1')^{-1} + K_t^* \|_E^{-1} \|_E.
\]

For large \( t \), \( K_t^* \) dominates \((T_1 S_0 T_1')\). Thus,

\[
\| S_t \|_E < \| \Phi_{2,t} \|_E^2 \| (K_t^*)^{-1} \|_E.
\]

Assuming the elements of \( \Phi_{2,t} \) are bounded, then

\[
(2.2.10) \quad \| S_t \|_E \to 0 \quad \text{as} \quad t \to \infty
\]

since the elements of \((K_t^*)^{-1}\) are approaching zero. This implies that the estimation error covariance matrix \( S_t \) converges to the null matrix as \( t \) increases.

Two implications from this discussion apply to the stochastic linear model with a nonrandom state vector.
They are

i) The added corrections to the state vector estimate at each recursion

\[ |\hat{\beta}_t - \hat{\beta}_{t-1}| \]

are damped out at the same rate as \( t^{-3/2} \) approaches zero, and

ii) The estimated state vector \( \hat{\beta}_t \) converges in probability to the true state vector \( \beta_t \) at the same rate as \( t^{-1/2} \) approaches zero.\(^{13}\)

The preceding discussion does nothing more than prove the consistency of the Bayes estimate in the nonrandom state vector model. The usefulness of this discussion is that it provides the essential arguments for the development of a lower bound for the covariance matrix \( S_t \) in the completely random model.

To study the large sample properties of the completely random Kalman model, our set of equations is

**OBS. EQ.** \( Y_t = X_t \beta_t + u_t \), and

**STATE EQ.** \( \beta_t = T_{t-1} \beta_{t-1} + V_t \)

\(^{12}\) For a proof of these two implications, see appendix 3.

\(^{13}\) Convergence in probability of a sequence of matrices \( M_n \) to a matrix \( N \) implies that for any \( \epsilon > 0 \),

\[
\lim_{n \to \infty} \Pr\{ |M_{n_{i,j}} - N_{i,j}| < \epsilon, \text{ for all } i,j \} = 1.
\]
where $u_t$ is $N(0, R_t)$, $v_t$ is $N(0, Q_t)$ and the vectors $u_t$ and $v_t$ are independent. The initial state vector $\beta_0$ is assumed to be random, independent of the errors $u_t$ and $v_t$ for all $t$, and possessing covariance matrix $S_0$.

We shall derive a lower bound for the covariance matrix $S_t$ for the Kalman model and compare it with that for the nonrandom state vector model. The purpose of the discussion is to show that under the model with a stochastically varying state vector the estimated state vector does not converge to the true state vector. Because of the presence of the state system excitation vector $v_t$ at each time point, it is not possible for us to ever explicitly determine $\beta_t$ from either $Y_1, Y_2, ..., Y_t$ or knowledge of another state vector $\beta_\ell$ for $\ell \neq t$. Hence, the covariance matrix $S_t$ in the Kalman model does not converge to the null matrix as the number of recursions increase but does so only if all the state system excitation vectors $v_t$ are null. This discussion is not intended to be practical in the sense of finding a computable lower bound for $S_t$ in the Kalman model. What we are attempting to show is that this lower bound does not converge to the null matrix as our recursions increase.

Suppose we write our set of equations for the Kalman model as

$$Y_t = Y_t^{(1)} + Y_t^{(2)}$$
and

\[ \hat{\beta}_t = \hat{\beta}_t^{(1)} + \hat{\beta}_t^{(2)} \]

where

\[ \hat{\gamma}_t^{(1)} = X_t \hat{\beta}_t^{(1)}, \]

\[ \hat{\beta}_t^{(1)} = T_t \hat{\beta}_t^{(1)} + \nu_t, \quad \hat{\beta}_0^{(1)} = 0 \]

and

\[ \hat{\gamma}_t^{(2)} = X_t \hat{\beta}_t^{(2)} + \nu_t \]

\[ \hat{\beta}_t^{(2)} = T_t \hat{\beta}_t^{(2)}, \quad \hat{\beta}_0^{(2)} = \beta_0. \]

The least squares estimator for the state vector of each subsystem is

\[ \hat{\beta}_t^{(i)} = \hat{\beta}_t^{(i)}|_{t-1} + M_t^{(i)} (\hat{\gamma}_t^{(i)} - X_t \hat{\beta}_t^{(i)}|_{t-1}), \quad i = 1, 2. \]

Thus, the estimator

\[ \hat{\beta}_t = \hat{\beta}_t^{(1)} + \hat{\beta}_t^{(2)} \]

is optimal if

\[ M_t^{(1)} = M_t^{(2)} = S_t X_t' R_t^{-1}, \]

the optimal smoothing matrix for the Kalman system (2.2.1).

The error covariance matrix for the estimated state vector is

\[ S_t = E(\hat{\beta}_t - \beta_t)(\hat{\beta}_t - \beta_t)' \]

(2.2.11) \[ = S_t^{(1)} + S_t^{(2)}, \]
where \( S^{(i)}_{t} = E(\hat{\beta}^{(i)}_{t} - \beta^{(i)}_{t})(\hat{\beta}^{(i)}_{t} - \beta^{(i)}_{t})\), \( i = 1, 2 \).

There is no cross product term in (2.2.11) since we chose \( \beta^{(1)}_{0} = 0 \).

**Subsystem without measurement errors**

To examine \( S^{(1)}_{t} \), note that we may write

\[
\hat{\beta}^{(1)}_{t} = T_{t}\hat{\beta}^{(1)}_{t-1} + M^{(1)}_{t}(X^{(1)}_{t} - X^{(1)}_{t}T_{t}\hat{\beta}^{(1)}_{t-1})
\]

\[
= T_{t}\hat{\beta}^{(1)}_{t-1} + M^{(1)}_{t}(X^{(1)}_{t}T_{t}\hat{\beta}^{(1)}_{t-1} + v_{t}) - X^{(1)}_{t}T_{t}\hat{\beta}^{(1)}_{t-1}
\]

\[
= (I-M^{(1)}_{t}X^{(1)}_{t})T_{t}\hat{\beta}^{(1)}_{t-1} + M^{(1)}_{t}X^{(1)}_{t}(T_{t}\hat{\beta}^{(1)}_{t-1} + v_{t}).
\]

The estimation error is then

(2.2.12) \( \beta^{(1)}_{t} - \hat{\beta}^{(1)}_{t} = (I-M^{(1)}_{t}X^{(1)}_{t})[T_{t}(\beta^{(1)}_{t-1} - \hat{\beta}^{(1)}_{t-1}) + v_{t}] \).

Suppose we define \( P_{i,i} = I \) and for \( i \neq j \)

\( P^{(1)}_{i,j} = (I-M^{(1)}_{i}X_{i})T_{i} \cdots (I-M^{(1)}_{j+1}X_{j+1})T_{j+1} \).

Expanding expression (2.3.6), we can write

(2.2.13) \( \beta^{(1)}_{t} - \hat{\beta}^{(1)}_{t} = P^{(1)}_{t,0} \beta^{(1)}_{0} - \hat{\beta}^{(1)}_{0} + \sum_{i=1}^{t} P^{(1)}_{t,i}v_{i} \).

Using the fact that \( \beta^{(1)}_{0} \) and hence \( \hat{\beta}^{(1)}_{0} = 0 \),

\[
S^{(1)}_{t} = \sum_{i=1}^{t} P^{(1)}_{t,i}Q_{i}P^{(1)}_{t,i}
\]
gives an expression for the covariance matrix of the first subsystem where \( M_t^{(1)} = S_t X_t R_t^{-1} \), the gain for the Kalman model.

**Subsystem with a nonrandom state vector**

For a nonrandom state system model, expression (2.2.8) indicates that the estimation error covariance matrix is

\[
S^*_t(2) = \{\phi_{t,2}(T_1 S_0 T_1')^{-1} \phi_{t,2} + \sum_{i=1}^{t} H_{t,i} R_i^{-1} H_{t,i}'\}^{-1}.
\]

\( S^*_t(2) \) is the covariance matrix for the system based on the employment of the smoothing matrix

\[
M_t^{(2)} = S_t^* X_t R_t^{-1}.
\]

Thus, if the second subsystem were based on the gain for the Kalman model, the resultant estimation error covariance matrix would satisfy

\[
S^*_t(2) < S_t(2)
\]

where \( S_t(2) \) is the covariance matrix based on the Kalman gain.

This is true because of the presence of the state system covariance matrices \( Q_t \) within the smoothing matrix \( M_t = S_t X_t R_t^{-1} \).

We know from (2.2.10) that \( S_t^*(2) \) converges to the null matrix as the number of recursions increase, but this condition does not hold for \( S_t(2) \).

A lower bound for \( S_t \) in the Kalman model

Since the covariance matrix

\[
S_t = S_t^{(1)} + S_t^{(2)}
\]
when we use the Kalman smoothing matrix $M_t = S_t X_t R_t^{-1}$, then if we use $M_t^{(2)} = S_t^{*} X_t R_t^{-1}$ as the smoothing matrix for the second subsystem, we have

$$S_t \geq S_t^{(1)} + S_t^{*}(2)$$

Thus, the covariance matrix provides a limiting lower bound for the Kalman covariance matrix $S_t$. That is,

$$\lim_{t \to \infty} S_t \geq \lim_{t \to \infty} \{S_t^{(1)} + S_t^{*}(2)\}$$

$$= \sum_{i=1}^{t} P_{t,i}^{(1)} Q_i P_{t,i}^{(1)'} + \lim_{t \to \infty} S_t^{*}(2)$$

$$= \sum_{i=1}^{t} P_{t,i}^{(1)} Q_i P_{t,i}^{(1)'}$$

Therefore, $S_t$ is bounded below by a function of all the non-null state vector excitation covariance matrices $Q_i$. If all the $Q_i$ are null, our model has a nonrandom state vector and $S_t$ converges to the null matrix.

If we are interested in the small sample behavior of $S_t$ and not its asymptotic behavior, then we could use $S_t^{*}(2)$ as a computable lower bound. Hence, at any time point $t$,

$$S_t \geq \{\phi, t, 2 (T_t S_t T_t')^{-1} \phi, t', 2 + \sum_{i=1}^{t} H_t, i' R_i H_t, i\}^{-1}$$

provides a lower bound for $S_t$. Equality results when all $Q_i$, $i = 1, 2, \ldots, t$ are null.

An upper bound for $S_t$ in the Kalman model

To find an upper bound for $S_t$, suppose we use as the
smoothing matrix for our second subsystem

\[ M^{(2)}_t = (X_t' R_t^{-1} X_t)^{-1} X_t' R_t^{-1}. \]

The estimation error for the second subsystem is

\[ \hat{\beta}^{(2)}_t - \bar{\beta}^{(2)}_t = (I - M^{(2)}_t X_t) [T_t (\hat{\beta}^{(2)}_t - \bar{\beta}^{(2)}_t)] - M^{(2)}_t U_t \]

(2.2.14)

\[ = P^{(2)}_{t,0} (\hat{p}_0 - \bar{p}_0) - \sum_{i=1}^{t} P^{(2)}_{t,i} M^{(2)}_i U_i \]

where \( P^{(2)}_{i,j} \) is defined like \( P^{(1)}_{i,j} \) except for \( M^{(2)}_i \) replacing \( M^{(1)}_i \). The error covariance matrix of the expression (2.2.14) is

\[ S^{(2)}_t = P^{(2)}_{t,0} S^{(2)}_0 P^{(2)}_{t,0}' + \sum_{i=1}^{t} P^{(2)}_{t,i} M^{(2)}_i R_i M^{(2)}_i P^{(2)}_{t,i}' \]

or, by substituting the value of \( M^{(2)}_t \),

(2.2.15) \[ \hat{S}^{(2)}_t = P^{(2)}_{t,0} S^{(2)}_0 P^{(2)}_{t,0}' + \sum_{i=1}^{t} P^{(2)}_{t,i} (X_i' R_i^{-1} X_i)^{-1} P^{(2)}_{t,i}'. \]

Expression (2.2.15) describes the covariance matrix of the estimation error for a Kalman model with state excitation vector \( v_t = 0 \), for all \( t \) and using \( M^{(2)}_t \) as the smoothing matrix. If we revise the second subsystem to include a variable state vector, then

(2.2.16) \[ S^*_t = \hat{S}^{(2)}_t + \sum_{i=1}^{t} P^{(2)}_{t,i} Q_i P^{(2)}_{t,i}' \]

is an estimate for \( S_t \) which is greater than \( S_t \) since \( S^*_t \) is based on the nonoptimal smoothing matrix \( M^{(2)}_t \) as
compared to the optimal smoothing matrix
\[ M_t = S_t X_t' R_t^{-1} \]
for this system.

The estimation error covariance matrix \( S_t \) is then limited by the inequality
\[ \sum_{i=1}^{t} P_{t,i} Q_i P_{t,i}' < S_t < S^*_t. \]

The upper bound and the asymptotic lower bound are more academic than useful since they depend on the (almost) unestimable excitation covariance matrices \( Q_i \). But a discussion of the lack of consistency of the estimated state vector elements in the Kalman model is necessary to motivate the test for detecting the presence of the Kalman model derived in section 3.3.2.

The implications of this discussion as they relate to the Kalman model are:

i) The added correction to the state vector at each point in time 
\[ |\hat{\beta}_t - \hat{\beta}_{t-1} | \]
does not necessarily damp out as \( t \) increases, and

ii) Since \( S_t \) does not approach the null matrix for increasing \( t \), the estimation error 
\[ |\hat{\beta}_t - \hat{\beta}_{t-1} | \]
does not necessarily converge to zero.

Furthermore, we have shown that in the random state vector Kalman model, because we do not have independent identically distributed random variables, the state vector estimate is not consistent in the sense of
\[ S_t = E(\hat{\beta}_t - \beta_t)(\hat{\beta}_t - \beta_t)' \]

converging to the null matrix for increasing \( t \). This lack of consistency exists even though appendix 2 shows that \( \hat{\beta}_t \) as defined by (2.2.4) is the Bayes estimate for \( \beta_t \) with the Kalman model.

2.3 Covariance Matrix Estimation in the Recursive Model

The discussion relative to the recursive model has assumed complete knowledge of the observation error and state vector error covariance matrices. When they are unknown, the iterative techniques of chapter 1 together with a weighting scheme is suggested for obtaining their estimated values.

2.3.1 Covariance Estimation When the Parameter Vector is Nonrandom or Constant

For the linear model

\[ (2.3.1) \quad Y_t = X_t\beta_t + \varepsilon_t \]

where \( \varepsilon_t \) is \( (0, \mathbb{I}_n) \), \( \mathbb{I}_n \) is of full rank, the matrix \( X_t \) is of rank \( p \), and

\[ Y_t: (nx1), \quad \beta_t: (px1), \]

suppose we have one observation vector \( Y \) at each of \( T \) time points. Furthermore, suppose that the covariance
matrix $\Phi_t$ is variable over these $T$ time points. It is assumed that $n > p$ and that

$$\beta_t = T_t \beta_{t-1}$$

where the transition matrix $T_t$ is known.

From chapter 1, to obtain an estimator for $\Phi_1$ given only $Y_1$, one applies the appropriate iterative scheme depending on the form of $\Phi_1$. After a predetermined number of iterations, the estimator $\hat{\Phi}_1$ is offered as the interactively smoothed covariance matrix estimator.

At the next recursion, our estimator $\hat{\Phi}_2$ should reflect not only the new observational data $Y_2$ but also the values of $\hat{\Phi}_1$ obtained in the previous step. We can accomplish this by allowing the initial weight matrix in the iterative process to be $\Phi_1$. Thus, the covariance matrix estimator at the $j$th recursive cycle is a smoothed estimator of all the available data $Y_1, Y_2, \ldots, Y_j$ up to that time.

The recursive estimation procedure may be summarized as follows:

Step 1. From an initial data set $(Y_1, X_1)$ apply the appropriate iterative scheme to the calculated residuals to obtain the estimated error covariance matrix $\hat{\Phi}_1$. This process is initiated by the choice of $\omega = I_n$ as the initial weight matrix.
Step 2. Using \( \hat{\Phi}_1 \) as the initial weight matrix, iteratively estimate \( \hat{\Phi}_2 \) from the second data set \((Y_2, X_2)\).

Step 3. In a recursive manner, initiate the iterations at the \( j \)th recursive cycle by the choice of \( \hat{\Phi}_{j-1} \) as initial weight matrix. Then incorporate the \( j \)th data set into the iterative scheme to obtain the estimator \( \hat{\Phi}_j \).

Step 3 is repeated at each recursive cycle for which we have a data set \((Y, X)\).

2.3.2 Covariance Matrix Estimation in the Kalman Model

The Kalman model, referenced as (2.2.1), assumes the observation errors and state system excitations are \( N(0, R_t) \) and \( N(0, Q_t) \), respectively. Here we have two covariance matrices to estimate at each recursive cycle. Furthermore, the actual excitations generated by the \( N(0, Q_t) \) process are unestimable, complicating the problem of estimating the matrix \( Q_t \).

Duncan's theorem allows us to estimate \( R_t \) and \( Q_t \) simultaneously by viewing the Kalman model as a partitioned general linear model. By (2.2.6), the Kalman model may be written in the form

\[
\begin{bmatrix}
\frac{Y_t}{\hat{\Phi}_t|_{t-1}}
\end{bmatrix}
= \begin{bmatrix}
X_t
\end{bmatrix} \begin{bmatrix}
\beta_t
\end{bmatrix} + \begin{bmatrix}
u_t
\end{bmatrix}
\]

\[
\begin{bmatrix}
\frac{Y_t}{\hat{\Phi}_t|_{t-1}}
\end{bmatrix}
\begin{bmatrix}
\frac{X_t}{\beta_t}
\end{bmatrix} = \begin{bmatrix}
I
\end{bmatrix} + \begin{bmatrix}
\xi_t
\end{bmatrix}
\]
where $\xi_t = T_t (\hat{\beta}_t - \beta_{t-1}) + Y_t$. The above partitioned model has an assumed fixed state vector even though $\beta_t$ in the Kalman model is random. Our working model is then of the form

\[(2.3.2) \quad Y_{pt} = X_{pt} \beta_t + \xi_{pt}\]

where

$$Y_{pt} \begin{bmatrix} Y_t \\ \hat{\beta}_t | t-1 \end{bmatrix}, \quad X_{pt} = \begin{bmatrix} X_t \\ I \end{bmatrix}, \quad \xi_{pt} = \begin{bmatrix} u_t \\ \xi_t \end{bmatrix}$$

and $\xi_{pt}$ is $\mathcal{N}(0, \Sigma_{pt})$. Thus, the estimation techniques of the previous section may be applied to (2.3.2) to find the covariance matrix estimator

$$\Sigma_{pt} = \begin{bmatrix} \hat{R}_t & 0 \\ 0 & \hat{S}_t | t-1 \end{bmatrix}$$

at each recursive cycle. At the first recursive cycle, the recursive process must be initiated by assumed initial values for $\hat{\beta}_0$ and $S_1 | 0$.

From the covariance matrix estimator $\Sigma_{pt}$ calculated at each recursion, we obtain
\( \hat{R}_t = \) the upper left \((mxn)\) submatrix of \( \hat{\mathbf{A}}_{pt} \), and

\[ \hat{Q}_t = \hat{S}_{t|t-1} - T_t \hat{S}_{t-1} T_t' \]

where \( \hat{S}_{t|t-1} \) is the lower right \((pxp)\) submatrix of \( \hat{\mathbf{A}}_{pt} \).

The estimation error covariance matrices \( S_t \) are estimated at each recursion by

\[ \hat{S}_t = [\hat{S}_{t|t-1} + X_t' R_t^{-1} X_t]^{-1}. \]
3.0 Experimental Studies

In this section we discuss simulation studies of three variance estimation schemes. The first study demonstrates the ability of the iterative estimation approach to estimate the observation variances of a linear regression model given only one set of observations. Next, the iterative technique is applied for estimating the variances at each time point of a linear model sequenced over time. The last study explores variance estimation in the Kalman recursive model.

In all studies a "true" process is generated with known regression parameters and observation and, when appropriate, state system variances. Then, by applying the variance estimation techniques of Chapter 1, variance estimates are used as weights to provide weighted estimates of the assumed unknown regression parameters. Accuracy of estimation is measured by the Euclidean norm of the difference between the estimator and its true value. It will be noted with emphasis that in the model

\[ Y = X\hat{\beta} + \epsilon \]

where \( \epsilon \) is \((0, \Sigma)\), \( Y: (nx1) \), \( \beta: (px1) \), the residuals \((Y - X\hat{\beta}(\omega))\) are associated with only \((n-p)\) degrees of freedom. Thus, if we attempt to estimate the elements of \( \Sigma \) from a function of the residuals, we can expect at most \((n-p)\) of the estimated variances to accurately reflect their true values due to the dependencies among the residuals.
3.1 Estimation of the Covariance Matrix \( \Sigma \) (Diagonal) From One Observation Vector

Assume that the linear model has independent errors and that \( Y : (7 \times 1) \) and \( \beta : (3 \times 1) \). For a known set of error variances, residuals are generated by a Gaussian (normal) generating process. These true residuals are entered into the model from which estimated residuals are obtained. Then the estimated residuals are incorporated into the iterative estimation schemes of Chapter 1 in an attempt to estimate their true, underlying variances. Two estimation techniques are employed; the diagonal approximation method (DAM) of section 1.3.4 and the general smoothing method (GSM) of section 1.3.2.

The Euclidean norm of the difference is used to measure the closeness of an estimated matrix to the true matrix being estimated. This norm is considered superior to the difference or determinant of the difference since the latter measures do not provide a sufficient condition for convergence in probability of the difference matrix. The norm giving the largest element of the difference matrix does provide a sufficient condition for convergence in probability but is considered herein to be inferior to the Euclidean norm since it ignores all other elements of the difference matrix but the largest element. Watson (19)
measures the closeness of two matrices by computing the ratio of their determinants. One can see that it is a simple task to construct numerous matrices, element by element quite different from each other, but all having the same determinant. Also, the Watson measure is applicable only for comparing square matrices of full rank where the Euclidean norm can be used as a measure on any matrix.

3.1.1 Choice of the Smoothing Constant

For a (7x3) model with a diagonal covariance matrix and true variances

\[ \begin{bmatrix} \hat{\Sigma} \end{bmatrix} = \text{diag}\{1,2,3,4,5,6,7\} \]

Table 3.1 lists the average value for the Euclidean norm \( \| \hat{\Sigma} - \Sigma \|_E \) at the first five iterations for 25 observation vectors when the exponential smoothing constant is allowed to vary. The GSM is employed for obtaining the values.

This presentation shows that for \( \alpha < 0.55 \), there is an apparent convergence of the norm for the first five iterations. As \( \alpha \) increases, the norm decreases at an earlier iteration. This is due to the construction of the smoothing model which weights with weight \( \alpha \) the new residual cross product information at each iterative cycle. If we are to restrict our analysis to five iterations within each recursion of a recursive linear model, a smoothing constant \( \alpha = 0.50 \) appears the best. However, a constant
which places increasing weight on the previous estimate, such as $\alpha = 1/N$, appears to be best for any number of iterations from one to five.

The preceding conclusions have been substantiated by further studies all employing the estimation technique referenced as (1.3.6). In each study the covariance matrix norm attained a minimum within five iterations. When for some choices of $\alpha > 0.5$ the norm tends to increase for succeeding iterations, it does not increase unbounded but tends to oscillate around its minimum possible value and eventually damp out.

Much more work needs to be done to find the best smoothing constant for the iterative schemes DAM and GSM. Specifying $\alpha = 0.5$ or $\alpha = 1/N$ as the best constants to use under any conditions is over simplifying the problem. More specifically, the best smoothing constant is a function of the design matrix $X$ and the unknown covariance matrix $\Sigma$. Further studies will explore the estimation of the proper smoothing constant more tailored for the specific problem to which it would be applied than the intuitively appealing choices $\alpha = 0.5$ or $\alpha = 1/N$. 
TABLE 3.1

Iteration

<table>
<thead>
<tr>
<th>Smoothing Constant</th>
<th>N = 1</th>
<th>N = 2</th>
<th>N = 3</th>
<th>N = 4</th>
<th>N = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha=0.2$</td>
<td>12.612</td>
<td>11.163</td>
<td>9.886</td>
<td>8.778</td>
<td>7.835</td>
</tr>
<tr>
<td>$\alpha=0.25$</td>
<td>12.210</td>
<td>10.462</td>
<td>8.987</td>
<td>7.775</td>
<td>6.816</td>
</tr>
<tr>
<td>$\alpha=0.3$</td>
<td>11.809</td>
<td>9.790</td>
<td>8.169</td>
<td>6.922</td>
<td>6.023</td>
</tr>
<tr>
<td>$\alpha=0.35$</td>
<td>11.409</td>
<td>9.149</td>
<td>7.433</td>
<td>6.217</td>
<td>5.441</td>
</tr>
<tr>
<td>$\alpha=0.4$</td>
<td>11.009</td>
<td>8.541</td>
<td>6.782</td>
<td>5.654</td>
<td>5.047</td>
</tr>
<tr>
<td>$\alpha=0.45$</td>
<td>10.611</td>
<td>7.966</td>
<td>6.216</td>
<td>5.227</td>
<td>4.815</td>
</tr>
<tr>
<td>$\alpha=0.5$</td>
<td>10.214</td>
<td>7.427</td>
<td>5.736</td>
<td>4.924</td>
<td>4.716</td>
</tr>
<tr>
<td>$\alpha=0.55$</td>
<td>9.817</td>
<td>6.925</td>
<td>5.340</td>
<td>4.733</td>
<td>4.722</td>
</tr>
<tr>
<td>$\alpha=0.6$</td>
<td>9.422</td>
<td>6.463</td>
<td>5.028</td>
<td>4.639</td>
<td>4.807</td>
</tr>
<tr>
<td>$\alpha=0.65$</td>
<td>9.028</td>
<td>6.044</td>
<td>4.796</td>
<td>4.628</td>
<td>4.945</td>
</tr>
<tr>
<td>$\alpha=0.7$</td>
<td>8.636</td>
<td>5.670</td>
<td>4.640</td>
<td>4.684</td>
<td>5.129</td>
</tr>
<tr>
<td>$\alpha=0.75$</td>
<td>8.245</td>
<td>5.343</td>
<td>4.556</td>
<td>4.793</td>
<td>5.330</td>
</tr>
<tr>
<td>$\alpha=0.8$</td>
<td>7.855</td>
<td>5.069</td>
<td>4.536</td>
<td>4.941</td>
<td>5.542</td>
</tr>
<tr>
<td>$\alpha=1/N$</td>
<td>5.367</td>
<td>5.015</td>
<td>4.824</td>
<td>4.778</td>
<td>4.649</td>
</tr>
</tbody>
</table>
3.1.2 Distribution of the Iterative Variance Estimators

Tables 3.2 and 3.3 list the means and the 95% confidence intervals for the mean from 25 iterative estimators on each of two different covariance matrices using four different estimation schemes. Each estimator is the iterative variance estimate from the fifth iterative cycle on a new set of generated observations. All iterations are initiated by letting the initial weights be unity. For each model, the DAM and the GSM with smoothing constants \( \alpha = 0.5 \) and \( \alpha = 1/N \), where \( N \) is the number of the iterative step, are employed for finding the variance estimates. The confidence intervals are constructed under assumptions of normality of the variance estimates, a questionable assumption in this case. The model and its variance estimates displayed within Table 3.2 has an assumed covariance matrix with all the true variances equal to 2.0. In Table 3.3, the underlying model has its true variances increasing from 1.0 to 7.0.

One can note a marked tendency for the DAM to underestimate the true variances of each model. This is due to the fact that the DAM is an approximate approach, returning estimators for the observation variances which are unbiased only if the weights used to obtain the computed residuals are the true variances of the model. Since our initial weights are unity which is at least as small
as any of the true variances in either model, the DAM, using these wrong weights as if they were true, will underestimate the variances.

On the other hand, the more general GSM is not an approximate approach when the true variances are not used as weights and appears to return, on the average, very accurate estimates for the observation variances in both the constant variance and increasing variance model.

A study of the confidence intervals for the observation variances shows that, with only one exception, at a given value for the smoothing constant the GSM returns more stable estimators for the variances than does the DAM. Since a normal generating process for generating normally distributed random variables may sometimes produce extremely unlikely values, this implies that the GSM is doing a better job than the DAM in smoothing out the effects of the extreme residuals.
<table>
<thead>
<tr>
<th></th>
<th>DAM $\alpha=0.5$</th>
<th>DAM $\alpha=1/N$</th>
<th>GSM $\alpha=0.5$</th>
<th>GSM $\alpha=1/N$</th>
<th>TRUE VARIANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_1$</td>
<td>1.4746 (1.6772)</td>
<td>1.3454 (1.7054)</td>
<td>1.7958 (1.9676)</td>
<td>1.7117 (2.0245)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>1.8798 2.0654</td>
<td>2.1394 2.3373</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>1.7254 (2.2312)</td>
<td>1.4887 (1.9831)</td>
<td>1.6990 (1.8946)</td>
<td>1.6788 (1.9080)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.7370 2.4775</td>
<td>2.0902 2.1372</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_3$</td>
<td>1.4851 (1.7971)</td>
<td>1.1790 (1.6768)</td>
<td>1.8519 (2.0179)</td>
<td>1.4666 (1.8516)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.1091 2.1746</td>
<td>2.1839 2.2466</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_4$</td>
<td>1.7676 (2.0974)</td>
<td>1.7753 (2.1217)</td>
<td>1.7394 (1.9872)</td>
<td>1.8061 (2.1295)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.4272 2.4671</td>
<td>2.2350 2.4529</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_5$</td>
<td>1.5565 (1.8547)</td>
<td>1.5595 (1.8465)</td>
<td>1.7155 (1.9171)</td>
<td>1.8879 (2.1219)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.1529 2.1345</td>
<td>2.1187 2.3559</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_6$</td>
<td>1.3900 (1.7830)</td>
<td>1.4893 (1.9837)</td>
<td>1.0841 (1.9575)</td>
<td>1.5869 (1.9923)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>2.1760 2.4781</td>
<td>2.8309 2.3977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_7$</td>
<td>1.0893 (1.3825)</td>
<td>1.0191 (1.3297)</td>
<td>1.8678 (2.1202)</td>
<td>1.4821 (1.7103)</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>1.6757 1.6403</td>
<td>2.3726 1.9385</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The values within each cell are the means (in parentheses) and the 95% confidence limits for the mean obtained from 25 estimates for each observation variance.
<table>
<thead>
<tr>
<th></th>
<th>DAM $\alpha=0.5$</th>
<th>DAM $\alpha=1/N$</th>
<th>GSM $\alpha=0.5$</th>
<th>GSM $\alpha=1/N$</th>
<th>TRUE VARIANCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_1$</td>
<td>0.5224 (0.5636)</td>
<td>0.0470 (0.1272)</td>
<td>0.9657 (1.0007)</td>
<td>0.9316 (1.0014)</td>
<td>1.0</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>1.2112 (1.9022)</td>
<td>1.2486 (1.8058)</td>
<td>1.4956 (2.0994)</td>
<td>1.8111 (2.1989)</td>
<td>2.0</td>
</tr>
<tr>
<td>$\sigma^2_3$</td>
<td>1.8067 (2.2541)</td>
<td>1.6779 (1.5527)</td>
<td>2.4396 (2.8226)</td>
<td>2.1981 (2.9641)</td>
<td>3.0</td>
</tr>
<tr>
<td>$\sigma^2_4$</td>
<td>2.8260 (3.5258)</td>
<td>2.8863 (3.5801)</td>
<td>3.3784 (4.1038)</td>
<td>3.6079 (4.2077)</td>
<td>4.0</td>
</tr>
<tr>
<td>$\sigma^2_5$</td>
<td>3.1811 (3.7013)</td>
<td>2.6426 (3.7036)</td>
<td>4.4909 (5.0009)</td>
<td>3.9819 (5.0019)</td>
<td>5.0</td>
</tr>
<tr>
<td>$\sigma^2_6$</td>
<td>4.0699 (5.1417)</td>
<td>3.2073 (4.5091)</td>
<td>4.8194 (5.8488)</td>
<td>4.4526 (5.7316)</td>
<td>6.0</td>
</tr>
<tr>
<td>$\sigma^2_7$</td>
<td>5.1291 (6.5525)</td>
<td>4.0789 (6.1201)</td>
<td>5.6879 (6.9953)</td>
<td>5.1760 (6.9906)</td>
<td>7.0</td>
</tr>
</tbody>
</table>
3.1.3 Goodness of the Estimators

The best estimator was earlier defined as the one with minimal covariance matrix. The weighted least squares estimator $\hat{\beta}(\frac{1}{w})$ is best since its covariance matrix $\Psi(\frac{1}{w})$ is less than that for the least squares estimator $\hat{\beta}(\omega)$, where $\omega$ is any weight matrix other than $\frac{1}{w}$. If $\Psi(\omega)$, the covariance matrix of $\hat{\beta}(\omega)$, is close to $\Psi(\frac{1}{w})$, then the estimator $\hat{\beta}(\omega)$ based on $\omega$ will be called a good estimator of the regression parameter $\beta$.

Table 3.4 compares the best estimates $\hat{\beta}(\frac{1}{w})$ with the iterative estimates for two different observation vectors from a $(7 \times 3)$ linear regression model. The Euclidean norm of the difference is given as a measure of closeness for these and other comparisons. The DAM is applied to the first observation vector while the GSM is applied to the second.

Both studies show that the variances are estimated more closely by using a smoothing constant $\alpha = 0.5$. Also, the analyses based on smoothing constant $\alpha = 0.5$ produce estimators $\hat{\beta}(\frac{1}{w})$ and $\Psi(\frac{1}{w})$ closer to the best linear estimator $\hat{\beta}(\frac{1}{w})$ and its covariance matrix $\Psi(\frac{1}{w})$ than those based on $\alpha = 1/N$.

It must be noted that each case, our analysis is based on 7 weights (variances) which are estimated from 7 residuals associated with only 4 degrees of freedom. Therefore, complete accuracy in our estimation of the weights cannot be expected.
**TABLE 3.4**

<table>
<thead>
<tr>
<th></th>
<th>DAM $\alpha=0.5$</th>
<th>DAM $\alpha=1/N$</th>
<th>GSM $\alpha=0.5$</th>
<th>GSM $\alpha=1/N$</th>
<th>TRUE VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta(\mathcal{F})_1$</td>
<td>4.808</td>
<td>4.808</td>
<td>4.939</td>
<td>4.937</td>
<td>4.0</td>
</tr>
<tr>
<td>$\beta(\mathcal{F})_2$</td>
<td>-2.064</td>
<td>-2.064</td>
<td>-2.057</td>
<td>-2.057</td>
<td>-2.0</td>
</tr>
<tr>
<td>$\beta(\mathcal{F})_3$</td>
<td>2.968</td>
<td>2.968</td>
<td>2.925</td>
<td>2.925</td>
<td>3.0</td>
</tr>
<tr>
<td>$\beta(\mathcal{F})_1$</td>
<td>4.385</td>
<td>6.160</td>
<td>5.853</td>
<td>6.496</td>
<td>4.0</td>
</tr>
<tr>
<td>$\beta(\mathcal{F})_2$</td>
<td>-1.947</td>
<td>-2.037</td>
<td>-1.969</td>
<td>-2.006</td>
<td>-2.0</td>
</tr>
<tr>
<td>$\beta(\mathcal{F})_3$</td>
<td>2.898</td>
<td>2.623</td>
<td>2.556</td>
<td>2.475</td>
<td>3.0</td>
</tr>
<tr>
<td>$|\hat{\mathcal{F}} - \mathcal{F}|_E$</td>
<td>2.171</td>
<td>2.516</td>
<td>6.040</td>
<td>6.963</td>
<td>...</td>
</tr>
<tr>
<td>$|\Psi(\hat{\mathcal{F}}) - \Psi(\mathcal{F})|_E$</td>
<td>2.388</td>
<td>7.785</td>
<td>2.211</td>
<td>4.107</td>
<td>...</td>
</tr>
<tr>
<td>$|\hat{\beta}(\mathcal{F}) - \beta(\mathcal{F})|_E$</td>
<td>0.444</td>
<td>1.396</td>
<td>0.988</td>
<td>1.622</td>
<td>...</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_1$</td>
<td>5.294</td>
<td>6.607</td>
<td>3.355</td>
<td>3.706</td>
<td>6.0</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_2$</td>
<td>0.522</td>
<td>0.022</td>
<td>0.367</td>
<td>0.140</td>
<td>1.0</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_3$</td>
<td>8.265</td>
<td>10.147</td>
<td>8.160</td>
<td>9.351</td>
<td>9.0</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_4$</td>
<td>7.101</td>
<td>7.791</td>
<td>6.580</td>
<td>7.143</td>
<td>8.0</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_5$</td>
<td>9.276</td>
<td>10.511</td>
<td>5.072</td>
<td>5.431</td>
<td>10.0</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_6$</td>
<td>0.758</td>
<td>0.582</td>
<td>0.740</td>
<td>0.706</td>
<td>1.0</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_7$</td>
<td>0.519</td>
<td>0.213</td>
<td>0.586</td>
<td>0.484</td>
<td>2.0</td>
</tr>
</tbody>
</table>
3.2 Variance Estimation in the Recursive Model

A linear model of the form

$$Y_t = X_t \beta + \epsilon_t$$

where $Y_t$: (4x1), $\beta$: (2x1) and $\epsilon_t$ is $(0, \Sigma_t)$ with $\Sigma_t$ diagonal has been generated at each of 25 time points. The variances of the observations, the diagonal elements of $\Sigma_t$, are allowed to vary over time.

Initially the estimation process is begun by letting $\hat{\Sigma}_1 = I_7$. Then, applying the iterative techniques of Chapter 1 we obtain estimates for the elements of $\hat{\Sigma}_1$ from the first observation vector $Y_1$ and the residuals generated at that recursive step. For each succeeding cycle of the recursion process, the initial weight matrix in the iterative scheme is the final iterative estimate for $\hat{\Sigma}_1$ from the previous recursive cycle.

Two types of error structures are generated for the examples of this section, normally distributed errors and a "stable" distribution of errors. The disadvantage of using normally distributed errors, and thus normally distributed observations, is that outliers even though unlikely do enter the model and confuse the estimation of variances. The iterative estimation procedure at best finds the most likely variance of each residual. Thus, an analysis based wholly on normally distributed observations
may give misleading information about the usefulness of using iterative techniques for the estimation of variances.

As an alternative to normally distributed observations, residual errors are generated for which extremely unlikely values are excluded. The only requirement of these filtered, more stable input errors is that they have a mean of zero with a variance equal to what is should be at that recursion. In essence, the stable input errors are more uniform than normal. The analyses based on the more stable input should be quite reflective of the major characteristics of the iterative estimation methods. In addition, the uniform input errors do not violate any normality assumptions since none of the estimation procedures of Chapter 1 require normality.

3.2.1 The Diagonal Approximation Method

As explained in section 1.3.4, the diagonal approximation method offers approximately unbiased estimators for the observation variances at each iterative step. Within each recursive cycle, the variance estimates are iteratively smoothed using a smoothing constant equal to $1/N$, where $N$ is the iterative step. Thus, within each recursion, the smoothed estimator for $\hat{\sigma}^2$ at the $l$th iterative step is

$$\omega_{l+1} = \alpha(\hat{\sigma}^2) + (1-\alpha)\omega_l$$
where $\alpha = 1/N$ and $\hat{a}$ is defined in (1.3.12).

From figures 1, 2, and 3 we observe the response of the DAM estimates to a constant, an increasing and a cyclic variance. All appear to track well with an apparent lag one effect present. This is due to the choice of the initial covariance matrix estimator at each recursion as the final iterative estimate from the previous recursion.

When the constant observation variance is changed from 1 to 5, figure 3 shows us that our estimates for the first and third observation variances are much less accurate than in figure 2. This is an indication that the residuals are associated with only two degrees of freedom and, thus, we cannot expect to gain accurate estimates of more than two observation variances.

In figure 4, a jump is introduced at the thirteenth recursion increasing the first observation variance from 1 to 5. The estimates for variances 2 and 4 are also excited at the thirteenth recursive cycle indicating the dependence among the residuals at that recursion.

It should be noted that the variance estimates which appear in figures 1, 2, and 3 represent the last iterative estimate within that recursive cycle. For the DAM, five iterations are performed within each recursion.
3.2.2 The General Smoothing Method

Since the DAM is applicable only when the observations are assumed to be uncorrelated, the general smoothing method (GSM) referenced as (1.3.6) offers a more versatile estimation technique. Assume the model

\[ Y_t = X\beta + \varepsilon_t \]

where \( Y_t \) : (4x1), \( \beta \) : (2x1) and \( \varepsilon_t \) is \((0, \Sigma_t)\) has uncorrelated observations at each time point. Thus, \( \Sigma_t \) is diagonal. The GSM has been applied to this model so that we might check the estimation ability of the GSM against that of the DAM.

Figures 5 and 6 both use observations generated from a normal distribution, but figure 6 uses a smoothing constant of 0.5 while figure 5 is based on the constant \( 1/N \), where \( N \) is the iterative cycle. There appears to be little discernable difference between the estimation ability of GSM in figures 5 and 6 and DAM from figure 1 where all three studies are tracking the same, underlying true variances and all are based on normal input.

In figure 7, we note that GSM does a much poorer job overall in estimating the same underlying variances that are estimated by DAM in figure 2. Figure 7 shows a definite lag 1 effect in estimating the fourth observation variance and a marked inability to accurately estimate the
third variance. The latter difficulty is due to the aforementioned fact that the four residuals are associated with only two degrees of freedom.

The results displayed in figure 8 may be compared with those from figure 4. When the first observation variance jumps from 1 to 5, the second and fourth variance estimates are greatly excited under DAM. However, the GSM appears to have only its second variance estimate perturbed while smoothing out the effects of the residual excitations in the fourth observation variance.

Figures 9a, 9b, 9c, and 9d study the degree to which each type of variance pattern influences the estimation of other variances. Only in figure 9a do we have an indication of a marked degree of instability in the variance estimates caused by a movement of one of the true variances. Thus, a jump in value of one of the true variances causes more instability in the estimation of the other variances than other movements of value of the true variances. The GSM is quite responsive to graduate changes in value of the true variances but not so responsive to abrupt changes.

Most all the previously described studies reveal accurate estimates for the second and fourth observation variances. In this section, when inaccurate estimates appear due to the lack of independence among the residuals, they are revealed in the estimation of the first and third variances. Figure 10 is a study to see whether the
inaccurate estimates are due to the pattern of movement of the variance or its location within the model. The patterns of movement are the same in figure 10 as they are in figure 8 but are changed in position. We see that variances 2 and 4 are most accurately estimated in both figure 8 and figure 10. Thus, the ability of a computed residual to determine its true, underlying variance is more a function of its position in the model than the movement of its true variance. This is another example of how a lack of degrees of freedom among the residuals causes dependencies among these residuals. In a (4x2) regression model, the two dependencies among the residuals make it impossible to expect accurate estimates of observation variances 1 and 3.

In figure 11, the objective has been to investigate the ability of the GSM to estimate the covariance matrix of the observational errors when some of the variances are very large. The second variance is allowed to increase 5 units per time period from 5 to 125, the fourth variance is a cosine function ranging between 35 and 45, while the first and third variances remain as in previous studies. The GSM was chosen instead of the DAM since figure 8 showed the GSM to be less effected by extreme changes in value of the true variances.

Observation variance 2 is estimated much less accurately than variance 4 since the changes in value of variance 2
are more abrupt than those of variance 4. However, we note that the estimates for the fourth observation variance don't catch up with the true values of variance 4 until the fifth recursion. This is because of the large difference between the starting values for the true and estimated values for the fourth observation variance. Overall, the estimated variances within the study referenced as figure 11 are quite responsive to the disproportionate differences between the first and third and the second and fourth variances.

For each of the studies within this section five iterations were performed within each recursion. As before, the initial variances within each recursion are the fifth iterative estimates from the previous recursion except at the first recursion where the initial choice $I = I_4$ is made.

3.2.3 Comparison of the Weighted Least Squares Estimators

For the studies which have a common set of true, underlying variances, table 3.5 compares their weighted least squares estimates with the best (minimum variance) estimates. A good estimator was defined in Chapter 1 as an estimator whose covariance matrix is "close" to the covariance matrix for the estimator based on the correct weight matrix. The Euclidean norm of the difference between the estimators based on the true and estimated weights
is given as a measure of closeness.

Given the linear model

\[ Y = X\beta + \varepsilon \]

where \( Y \): (4x1), \( \beta \): (2x1), and \( \varepsilon \) is \((0, \mathbb{I})\) table 3.4 lists the best weighted estimator

\[ \hat{\beta}(\mathbb{S}) = (X'\mathbb{S}^{-1}X)^{-1}X'\mathbb{S}^{-1}Y, \]

the weighted estimator based on an estimated error covariance matrix \( \mathbb{S} \)

\[ \hat{\beta}(\mathbb{S}) = (X'\mathbb{S}X)^{-1}X'\mathbb{S}Y, \]

the Euclidean norm of the difference

\[ \| \psi(\mathbb{S}) - \psi(\mathbb{S}) \|, \]

and the Euclidean norm of the difference

\[ \| \hat{\beta}(\mathbb{S}) - \hat{\beta}(\mathbb{S}) \|. \]

These are denoted as EN and ENB, respectively, within table 3.5.

The important item to note from this presentation is how near the weighted estimators are to the best estimators.

1. The estimator covariance matrices \( \psi(\mathbb{S}) \) and \( \psi(\hat{\mathbb{S}}) \) are defined in (1.1.2) and (1.1.1).

2. The Euclidean norm is defined in footnote 4 within Chapter 1.
As we mentioned earlier, we seek estimators with the smallest possible (least positive definite) covariance matrix. These estimators individually are not necessarily close to their true values, but collectively give rise to a covariance matrix nearly as small as the covariance matrix for the estimators based on the correct weights.

We can see the closeness of $\Psi(\mathbf{\hat{y}})$ to $\Psi(\mathbf{\hat{y}})$ and of $\hat{\mathbf{\beta}}(\mathbf{\hat{y}})$ to $\hat{\mathbf{\beta}}(\mathbf{\hat{y}})$ from the presentations of table 3.4. These listings also indicate that GSM is better than DAM by noting the sizes of the Euclidean norm values $EN$ and $ENB$. In the first table, the norm for GSM is less than that for DAM in 11 of 16 cases, while in the second table, the norm for GSM is the smallest in all 16 cases. This implies that the more general GSM is to be preferred over the DAM even though the true error covariance matrix is diagonal.
Figure 1

**Diagonal Approximation Method**

\[ \alpha = \frac{1}{N} \]

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]

\[ \sigma^2(3) \]

\[ \sigma^2(4) \]

**Legend:**

\( \sigma^2(i) \) = observation variance \( i \)

t = the recursive cycle

The points are the true variances

The crosses are the iterative variance estimates.
**Figure 2**

Diagonal Approximation Method               Stable Input

\[ \alpha = \frac{1}{N} \]

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]
Figure 2 (con't)
Figure 3

Diagonal Approximation Method

$$\alpha = 1/N$$

Stable Input
Figure 4 (con't)

$\sigma^2(3)$

$\sigma^2(4)$
Diagonal Approximation Method

Stable Input

$\alpha = \frac{1}{N}$

Figure 4
Figure 5

General Smoothing Method

\[ \alpha = \frac{1}{N} \]

\( \sigma^2(1) \)

\( \sigma^2(2) \)
Figure 5 (con't)

\[ \sigma^2(3) \]

\[ \sigma^2(4) \]
Figure 6

General Smoothing Method

\[ \alpha = 0.5 \]

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]

\[ \sigma^2(3) \]

\[ \sigma^2(4) \]
General Smoothing Method

$\alpha = 1/N$

Figure 7

Stable Input

$\sigma^2(1)$

$\sigma^2(2)$

$\sigma^2(3)$

$\sigma^2(4)$
General Smoothing Method

Stable Input

\[ \alpha = \frac{1}{N} \]

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]
Figure 8 (con't)

\[ \sigma^2(3) \]

\[ \sigma^2(4) \]
Figure 9(a)

**General Smoothing Method**

\[ \alpha = \frac{1}{N} \]

**Stable Input**

\[ \sigma^2(1) \]
\[ \sigma^2(2) \]
\[ \sigma^2(3) \]
\[ \sigma^2(4) \]
Figure 9(b)

**General Smoothing Method**

\[ a = \frac{1}{N} \]

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]

\[ \sigma^2(3) \]

\[ \sigma^2(4) \]

**Stable Input**
Figure 9(c)

General Smoothing Method

\[ \alpha = \frac{1}{N} \]

Stable Input
General Smoothing Method

\[ \alpha = \frac{1}{N} \]

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]

\[ \sigma^2(3) \]

\[ \sigma^2(4) \]
General Smoothing Method

\[ \alpha = \frac{1}{N} \]

Figure 10

Stable Input

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]
Figure 10 (con't)
Figure 11

**General Smoothing Approach**

\[ \alpha = \frac{1}{N} \]

Stable Input

\[ \sigma^2(1) \]

\[ \sigma^2(2) \]
## TABLE 3.5(a)
### Recursive Cycle

<table>
<thead>
<tr>
<th></th>
<th>$t = 3$</th>
<th>$t = 6$</th>
<th>$t = 9$</th>
<th>$t = 12$</th>
<th>$t = 15$</th>
<th>$t = 18$</th>
<th>$t = 21$</th>
<th>$t = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{\hat{\theta}_1}$</td>
<td>0.16437</td>
<td>-2.78256</td>
<td>-1.36128</td>
<td>-1.65856</td>
<td>1.69943</td>
<td>-2.54644</td>
<td>0.78657</td>
<td>-4.73520</td>
</tr>
<tr>
<td>$\hat{\beta}_{\hat{\phi}_1}$</td>
<td>1.66654</td>
<td>2.75529</td>
<td>2.11743</td>
<td>2.29338</td>
<td>0.77983</td>
<td>2.76777</td>
<td>1.10380</td>
<td>3.77306</td>
</tr>
<tr>
<td>$\hat{\beta}_{\hat{\phi}_2}$</td>
<td>0.59315</td>
<td>-3.49677</td>
<td>-1.84175</td>
<td>-2.22777</td>
<td>1.84064</td>
<td>-2.18902</td>
<td>0.31487</td>
<td>-4.66756</td>
</tr>
<tr>
<td>$\hat{\beta}_{\hat{\phi}_2}$</td>
<td>0.60556</td>
<td>-2.94584</td>
<td>-1.34558</td>
<td>-1.77194</td>
<td>2.01887</td>
<td>-2.58702</td>
<td>0.94501</td>
<td>-4.81848</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$t = 3$</th>
<th>$t = 6$</th>
<th>$t = 9$</th>
<th>$t = 12$</th>
<th>$t = 15$</th>
<th>$t = 18$</th>
<th>$t = 21$</th>
<th>$t = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{\hat{\theta}_1}$</td>
<td>1.46594</td>
<td>3.14204</td>
<td>2.42412</td>
<td>2.63209</td>
<td>0.74532</td>
<td>2.52700</td>
<td>1.31610</td>
<td>3.73731</td>
</tr>
<tr>
<td>$\hat{\beta}_{\hat{\phi}_1}$</td>
<td>1.46419</td>
<td>2.92368</td>
<td>2.12601</td>
<td>2.34526</td>
<td>0.66578</td>
<td>2.77036</td>
<td>1.08663</td>
<td>3.78425</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$t = 3$</th>
<th>$t = 6$</th>
<th>$t = 9$</th>
<th>$t = 12$</th>
<th>$t = 15$</th>
<th>$t = 18$</th>
<th>$t = 21$</th>
<th>$t = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>0.45413</td>
<td>9.13854</td>
<td>0.52382</td>
<td>0.27194</td>
<td>0.75436</td>
<td>0.24643</td>
<td>0.08083</td>
<td>1.11744</td>
</tr>
<tr>
<td></td>
<td>0.82662</td>
<td>2.53887</td>
<td>0.02943</td>
<td>0.02276</td>
<td>0.29954</td>
<td>0.00469</td>
<td>0.00957</td>
<td>3.48925</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$t = 3$</th>
<th>$t = 6$</th>
<th>$t = 9$</th>
<th>$t = 12$</th>
<th>$t = 15$</th>
<th>$t = 18$</th>
<th>$t = 21$</th>
<th>$t = 24$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENB</td>
<td>0.47337</td>
<td>0.81220</td>
<td>0.57001</td>
<td>0.66236</td>
<td>0.14536</td>
<td>0.43095</td>
<td>0.51728</td>
<td>0.07651</td>
</tr>
<tr>
<td></td>
<td>0.48538</td>
<td>0.23455</td>
<td>0.01789</td>
<td>0.12469</td>
<td>0.33918</td>
<td>0.04066</td>
<td>0.15936</td>
<td>0.08402</td>
</tr>
</tbody>
</table>

Figure Where Presented

- F-1
- F-5
TABLE 3.5(b)  
Recursive Cycle

<table>
<thead>
<tr>
<th></th>
<th>t = 3</th>
<th>t = 6</th>
<th>t = 9</th>
<th>t = 12</th>
<th>t = 15</th>
<th>t = 18</th>
<th>t = 21</th>
<th>t = 24</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}(\mathbb{S})_1$</td>
<td>-3.53336</td>
<td>1.17635</td>
<td>-3.39060</td>
<td>1.67869</td>
<td>-3.82086</td>
<td>1.44953</td>
<td>-3.58162</td>
<td>1.90819</td>
</tr>
<tr>
<td>$\hat{\beta}(\mathbb{S})_2$</td>
<td>3.17724</td>
<td>0.96478</td>
<td>3.12973</td>
<td>0.74839</td>
<td>3.40820</td>
<td>0.73494</td>
<td>3.32764</td>
<td>0.53754</td>
</tr>
<tr>
<td>$\hat{\beta}(\mathbb{S})_1$</td>
<td>-4.08566</td>
<td>1.61322</td>
<td>-3.70151</td>
<td>2.05218</td>
<td>-4.55150</td>
<td>2.08367</td>
<td>-4.05533</td>
<td>2.26271</td>
</tr>
<tr>
<td>$\hat{\beta}(\mathbb{S})_2$</td>
<td>3.43215</td>
<td>0.75466</td>
<td>3.28160</td>
<td>0.57810</td>
<td>3.85415</td>
<td>0.32665</td>
<td>3.66119</td>
<td>0.27187</td>
</tr>
<tr>
<td>$\hat{\beta}(\mathbb{S})_1$</td>
<td>-3.96741</td>
<td>1.44704</td>
<td>-3.56850</td>
<td>1.94660</td>
<td>-4.19373</td>
<td>1.77570</td>
<td>-3.79513</td>
<td>2.06519</td>
</tr>
<tr>
<td>$\hat{\beta}(\mathbb{S})_2$</td>
<td>3.37569</td>
<td>0.82825</td>
<td>3.22693</td>
<td>0.61716</td>
<td>3.66125</td>
<td>0.49502</td>
<td>3.51105</td>
<td>0.39860</td>
</tr>
<tr>
<td>EN</td>
<td>0.43915</td>
<td>0.19864</td>
<td>0.10866</td>
<td>0.15825</td>
<td>0.95817</td>
<td>0.70118</td>
<td>0.75492</td>
<td>1.96366</td>
</tr>
<tr>
<td>ENB</td>
<td>0.29338</td>
<td>0.11457</td>
<td>0.08573</td>
<td>0.14574</td>
<td>0.31092</td>
<td>0.21959</td>
<td>0.16237</td>
<td>0.18477</td>
</tr>
<tr>
<td>ENB</td>
<td>0.60829</td>
<td>0.48477</td>
<td>0.34602</td>
<td>0.41049</td>
<td>0.85598</td>
<td>0.75422</td>
<td>0.57936</td>
<td>0.44301</td>
</tr>
<tr>
<td>ENB</td>
<td>0.47726</td>
<td>0.30318</td>
<td>0.20272</td>
<td>0.29833</td>
<td>0.45063</td>
<td>0.40491</td>
<td>0.28147</td>
<td>0.20965</td>
</tr>
</tbody>
</table>

Figure Where Presented: F-4,8
3.2.4 Discussion of the Experimental Results

In the experimental studies shown in figures 1 through 11 of the previous section, we employed a (4x2) regression model with design matrix and parameter vector, respectively

\[
X = \begin{bmatrix}
2 & 5 \\
3 & 6 \\
1 & 3 \\
3 & 4 \\
\end{bmatrix}, \quad \text{and} \quad \beta = \begin{bmatrix}
-1 \\
2 \\
\end{bmatrix}.
\]

Many of the studies using a stable input distribution for the errors, such as those listed in figures 3, 4, 6 and 8, show a marked inability to accurately estimate the first and third variances. This is due partly to a problem of identifiability and partly to our choice of design matrix. The stable distribution is simply a truncated normal with mean zero and a variance equal to the variance of the true model at each recursion. It is more uniform than normal. When we combine such an array with our \(X\) matrix and \(\beta\) vector to find our observations

\[
Y = X\beta + \varepsilon
\]

we find that with the choice of \(X\) and \(\beta\) as above, \(y_1\) and \(y_2\) are nearly equal and \(y_3\) and \(y_4\) are equal. When we compute our residuals to estimate the unknown variances, again because of our choice of design matrix \(X\), our first and third residuals are much too small, tending to
underestimate the first and third variances. Only when some of
the true variances are very large as in figure 11 are the
residuals, based on the weight matrix which estimates these
variances from the previous recursion, not significantly
underestimating their true, underlying variances. This is
because the large variances tend to generate large random
errors for the model reducing the chance for confusion in
identifiability between $y_1$ and $y_2$ and between $y_3$ and $y_4$.

A separate set of studies were run to check the depen-
dence of the estimation technique on the design matrix.
When the first and second rows and the third and fourth rows
of the design matrix (3.2.1) were interchanged the second
and fourth variances were underestimated while the first and
third variances were accurately estimated. For varying choices
for the design matrix accurate estimates for all four variances
were obtained, an example being shown in figure 2. Thus, the
estimation techniques referenced as (1.3.6) and (1.3.11) are
not order dependent in terms of always giving inaccurate
estimates for the first and third variances in a (4x2) linear
regression model.

In practice, we cannot discover when the design matrix
will cause confusion since the parameter vector is unknown.
Furthermore, we know that the computed residuals on which our
variance estimates are based are associated with only \( n - p \) degrees of freedom. Thus, the computed residuals are not independent even though the true errors of the model are independent. The inaccuracies noted within the foregoing studies were a function of an unfortunate choice of design matrix and residuals. However, when we use normally distributed residuals tables 3.2 and 3.3 show that on the average the iterative variance estimates are very near the true values being estimated.

It should be noted that the studies herein are designed to illustrate an estimation technique and to provide a general algorithm for the estimation of variances in linear regression models. Further studies need to be undertaken to examine the aforementioned problems of identifiability and dependence upon the design matrix.
3.3 Variance Estimation in the Kalman Model

When the general linear regression model has a state vector with random components, the model becomes the Kalman model. As mentioned earlier, the model is of the form

\[ Y_t = X_t \beta_t + u_t \]

where

\[ \beta_t = T_t \beta_{t-1} + V_t \]

and where \( u_t \) is \( N(0, R_t) \), \( V_t \) is \( N(0, Q_t) \) and \( u_t, V_t \) are independent.

The following experiment is conducted on a model with four observations and two state parameters sequenced over 25 time points. The true variances of the observations and state excitations vary over time.

3.3.1 Variance Estimation

According to the discussions of appendix 4 and sections (2.2.2), the random parameter Kalman model may be written as a model with assumed fixed regression parameters as

\[
\begin{bmatrix}
\hat{Y}_t \\
\hat{\beta}_t |_{t-1}
\end{bmatrix}
= \begin{bmatrix}
X_t \\
I
\end{bmatrix}
\begin{bmatrix}
\beta_t \\
\xi_t
\end{bmatrix}
= \begin{bmatrix}
u_t \\
\xi_t
\end{bmatrix}
\]

(3.3.1) or \( \hat{Y}_{pt} = X_{pt} \beta_t + \varepsilon_{pt} \),
where

\[ \varepsilon_{pt} \sim N(0, \begin{bmatrix} R_t & 0 \\ 0 & S_{t|t-1} \end{bmatrix}) \]

Thus, by applying the recursive and iterative techniques to the assumed fixed model (3.3.1) we obtain estimates for the observation variances \( R_t \) and the estimation error covariance matrix \( S_{t|t-1} \). That is, at each time period, the weight matrix at the \( \ell \)th iterative step is

\[
\omega_{t,\ell} = \begin{bmatrix} \hat{R}_t & 0 \\ 0 & \hat{S}_{t|t-1} \end{bmatrix}
\]

where \( \hat{R}_t \) is the estimated observation variance matrix and \( \hat{S}_{t|t-1} \) is the estimated estimation error covariance matrix. The state excitation covariance matrix is estimated by

\[
\hat{Q}_t = \hat{S}_{t|t-1} - T_t \hat{S}_t \hat{T}_t'
\]

where

\[
\hat{S}_{t-1} = [\hat{S}_{t-2|t-1}^{-1} + X_t' \hat{R}_t^{-1} X_t]^{-1}
\]

Figure 12 displays the results of employing the general smoothing method (GSM) with smoothing constant
\[ \alpha = \frac{1}{N}, \text{ where } N \text{ is the number of the iterative step.} \]

Note here that by applying Duncan's Theorem to our original model, we are expanding a \((4 \times 2)\) model into a \((6 \times 2)\) model. Thus, the residuals from the revised model (3.3.1) are associated with four degrees of freedom.

Of the six variances, observation variances \(R(3,3)\) and \(R(4,4)\) and state excitation variance \(Q(2,2)\) seem to track least well. However, all of these variances are more closely estimated than some of the poorer estimates, such as the third observation variances from figures 7 and 8, in the recursive models with nonrandom state vectors. Thus, the additional degrees of freedom associated with the residuals of the Kalman model appear to increase the accuracy of our variance estimates.

The impressive note of figure 12 is the accuracy of the estimated state excitation error variances. Since the state vector excitation errors are completely unestimable, the estimation of their underlying variances would appear impossible. However, through the dependencies among the residuals and the smoothing of estimates, the GSM is able to obtain accurate estimates of these variances.

3.3.2 Detection of the Kalman Model

When in practice the data analyst chooses to use a linear model to represent his data, he must decide which type of linear model is appropriate. If his model is
restricted to one point in time, the data analyst need not be concerned about the specifics of the design. But if his estimation and inference is over a wide time spectrum, it is important to determine whether the model possesses random or nonrandom state parameters.

From (2.3.4), the Euclidean norm of the covariance matrix $S_t$ for the recursive model with nonrandom or constant state vector satisfies

$$||S_t||_E \rightarrow 0, \text{ as } t \rightarrow \infty.$$  

Thus, under the nonrandom or constant model

$$||S_1||_E > ||S_2||_E > \cdots > ||S_{t-1}||_E > ||S_t||_E.$$  

However, under the random state vector (Kalman) model no such relationship exists. From the discussions of section 2.2., the most we can say concerning the random model is that the Euclidean norm of $S_t$ is randomly distributed over time.

In practice, our estimator for $S_t$ is

$$\hat{S}_t = [\hat{S}_{t|t-1}^{-1} + X_t' \hat{R}_t^{-1} X_t]^{-1},$$

where the estimates $\hat{S}_{t|t-1}$ and $\hat{R}_t$ are taken from the weight matrix of the partitioned model (3.3.1). Because of the presence of sample and design errors, the Euclidean norms for the nonrandom model do not necessarily form a
monotonically decreasing series when we replace $S_t$ by $\hat{S}_t$. Nevertheless, we shall assume that the
\[
\lim_{t \to \infty} |\hat{S}_t|_E = 0
\]
in the presence of the nonrandom state vector model.

Under the hypothesis that the Kalman model is present,
\[
\Pr(|S_t|_E > |S_{t-1}|_E) = \Pr(|S_t|_E < |S_{t-1}|_E) = \frac{1}{2}
\]
for $t = 1,2,\ldots,T$. This implies that we could perform an exact binomial test and compute
\[
\Pr[z \geq K] = \sum_{t=k}^{T} \binom{T}{t} \left(\frac{1}{2}\right)^T
\]
where $K$ is the observed number of negative differences of the form

(3.3.2) \quad |\hat{S}_t|_E - |\hat{S}_{t-1}|_E

and $z$ is the binomial random variable.

Figure 13 plots the values of $|S_t|_E$ and $|\hat{S}_t|_E$ at each of 25 time points for the Kalman model and for the same recursive model with the state excitation variances $Q = 0$ at each recursion. For the Kalman model, we have 14 positive and 10 negative differences using both the Euclidean norm of the true and estimated estimation error covariance.
matrices. Applying the binomial test, the probability that we would obtain at least 10 negative differences of the form (3.3.2) in 24 trials with \( p = \frac{1}{2} \) is

\[
Pr[z \geq 10] = \sum_{t=10}^{24} \binom{24}{t} \left( \frac{1}{2} \right)^t = 0.73.
\]

Thus, it is very likely that under the Kalman model we would obtain as many as 10 negative differences of the form (3.3.2).

In the presence of the nonrandom state vector recursive model note how the norms \( ||S_t||_E \) are monotonically decreasing over time. But using the estimated matrices \( \hat{S}_t \), we have 6 positive changes of value and 18 negative changes of value of the norm \( ||\hat{S}_t||_E \). Under the hypothesis of the random model, \( p = \frac{1}{2} \) and

\[
Pr[z \geq 18] = \sum_{t=18}^{24} \binom{24}{t} \left( \frac{1}{2} \right)^t = 0.003.
\]

Therefore, it is very unlikely that the Kalman model is present.

The binomial test suggested herein is not very powerful when the number of recursions \( T \) is small. Since under the Kalman model we tend to get runs of moderate length where the signs of the differences (3.3.2) remain unchanged, the binomial test may be misleading when \( T \) is
small. However, if when \( T \) is small we begin with a run of positive differences of the form (3.3.2) we have very good evidence to assume that the Kalman model is present.

3.3.3 Dependence Upon Initial Values

In order to initiate the Kalman recursive system, one needs initial estimates for the state vector \( \hat{\beta}_0 \) and the estimation error covariance matrix \( S_0 \). Figure 14 shows the behavior of the state vector estimator \( \hat{\beta}_t \) when accurate and inaccurate initial estimates \( \hat{\beta}_0 \) are employed both for the completely random Kalman model and for the nonrandom recursive model with \( Q = 0 \) at each recursion. Behavior of \( \hat{\beta}_t \) is examined by measuring the closeness of \( \hat{\beta}_t \) to \( \beta_t \), the true value, by calculating the Euclidean norm

\[
(3.3.3) \quad ||\hat{\beta}_t - \beta_t||_E.
\]

In the completely random Kalman model the difference between the norms (3.3.3) for accurate and inaccurate initial state vectors damps out to a negligible amount after four recursions. However, when the nonrandom recursive model is present, the norm (3.3.3) obtained from an inaccurate initial estimate \( \hat{\beta}_0 \) appear slow to catch those calculated from having used accurate initial estimates. Thus, the effects of inaccurate estimates are damped out
Figure 12

Observation and State Variances for the Kalman Model

GSM

\( \alpha = \frac{1}{N} \)
Figure 12 (con't)
Figure 13
Behavior of the Estimation Error Covariance Matrix in the Kalman Model

Kalman model employing the true variances as weights

Legend:
Points represent model with variable state vector
Crosses represent model with nonrandom state vector
Figure 14

Behavior of $|\hat{\beta}_t - \beta_t|$ in the Recursive Model

Recursive model with nonrandom state vector

Recursive (Kalman) model with random state vector

Legend:
Points represent model with accurate initial state vector
Crosses represent model with inaccurate initial state vector
quickly in the completely random recursive model but are damped out very slowly at best in the nonrandom recursive model. This rather surprising result suggests that we need not worry about the initial estimate for $R_0$ when the completely random (Kalman) recursive model is present but that we do need to be concerned about its accuracy otherwise.

3.4 Conclusions

Very little appears in the literature concerning the iterative estimation of the variances of the observations of a linear model from one data set. Some authors discuss the iterative estimation of the parameter vector of a linear regression model from an assumed or derived set of weights where the interest is in estimation of the regression parameters and not of the weights. The studies herein assumed that since the best estimates of the regression parameters are obtained by using the true variances as the weights, if we can then obtain good estimates of the variances, the weighted parameter vector estimator will be closer to the best parameter vector estimator than any other estimator we could use. Thus, emphasis is placed on the good estimation of the observation variances and then, by using them as weights to obtain good estimates for the regression parameters.

Goodman (9) and Mallios (14) are the only two authors
who discuss an estimation approach similar to the iterative approach suggested within this paper. Both suggest an iterative approach using only the products of the residuals as the weight matrix for the estimation of the parameter vector at each iterative cycle. This author has found that such a procedure tends to explode since we are compounding our estimation errors in the residuals at each iterative cycle. A smoothed estimator for the weight matrix at each iterative cycle, combining the previous weight matrix and a function of the residual products, will guarantee convergence if the smoothing matrix contains no eigenvalue greater than 1 in absolute value.

According to Deutsch (6) the solution of the likelihood equation can be approximated by iteration techniques. Perhaps the iterative regression estimation techniques suggested herein are providing approximate solutions of the likelihood equation for the parameter vector $\beta$ and the error covariance matrix $\Sigma$. This author has not been able to satisfactorily answer this question for himself. Evidence in support of the conjecture that the iterative regression estimation techniques described herein provide solutions to the likelihood equations is given by tables 3.2 and 3.3 which show how closely the observation variances can be estimated by the iterative techniques.

The experimental studies of Chapter 3 suggest the
following conclusions in relation to the iterative and recursive estimation schemes discussed in Chapters 1 and 2:

i. The general smoothing method (GSM), as defined in section 1.3.2, offers the iterative scheme which is the most likely iterative scheme to produce parameter estimators nearest the best estimators when the error covariance matrix $\Sigma$ is diagonal;

ii. The choice of $\alpha = 0.5$ or $\alpha = 1/N$ are the smoothing constants which provide the most accurate estimates of the observation variances to be used within the iterative process;

iii. Three to five iterations per recursion are sufficient to obtain good variance estimates;

iv. The iterative estimation scheme, when employed for the estimation of the observation variances at each step of a recursive process, are quite responsive to gradual changes of the true variances. Abrupt changes cause estimation difficulties;

v. Even though the true model may possess independent errors, the computed residuals are dependent since they are associated with only $(n-p)$ degrees of freedom;

vi. A recursive model with random state components (Kalman) generates accurate estimates for the state parameters after about four recursions even though the initial state values used in the recursion may
be very inaccurate. For a nonrandom state vector model, many recursions are required;
.vii. The GSM may be used to obtain useful estimates of the state excitation variances in the Kalman model.

From the expression relating the computed residuals to the true error covariance matrix

\[ E(\mathbf{e} \mathbf{e}') = (I-K) \mathbf{S} (I-K)', \]  

the computational procedures of Chapter 1 have been derived in an attempt to estimate the covariance matrix \( \mathbf{S} \).

Other solution techniques which one could apply to expression (3.4.1) are:

i) Two stage least squares;

ii) A Seidel iterative procedure, much like the DAM;

iii) A generalized inverse approach;

iv) Triple exponential smoothing as defined and discussed by Robert G. Brown (4).

All these approaches were explored in separate studies but were found to be less desirable than those suggested in Chapter 1 for the estimation of the elements of \( \mathbf{S} \) in expression (3.4.1)
Appendix 1

Under assumptions of normality of the error vector, the prior estimator for the state vector of the recursive model is sufficient from all the past data for estimating the current state vector. This idea will be shown to be true for the Kalman model. Its truth for the general linear recursive model may be shown by performing the substitutions indicated by (2.2.5).

Definition
Suppose \( x_1, x_2, \ldots, x_n \) is a sample selected from a population having probability density function \( f(x|\theta) \). If \( \hat{\theta} = g(x_1, x_2, \ldots, x_n) \) is an estimator of \( \theta \) such that the conditional expectation \( E(\hat{\theta}|x_1, x_2, \ldots, x_n) \) does not depend on \( \theta \), then \( \hat{\theta} \) is a sufficient estimator for \( \theta \).

Theorem
Let our model be represented by the recursive equations

\[
Y_t = X_t \beta_t + u_t \\
\beta_t = T_t \beta_{t-1} + v_t, \quad t = 1, 2, \ldots
\]

where \( u_t \) is \( \mathcal{N}(0, R_t) \), \( v_t \) is \( \mathcal{N}(0, Q_t) \), \( E(u_t v_t') = 0 \), \( Y_t: (nx1), \beta_t: (px1) \), \( n > p \), and the rank of \( X_t \) is \( p \).

The prior estimator for the state vector \( \hat{\beta}_t \) is

\[
\hat{\beta}_t|t-1 = T_t \hat{\beta}_{t-1} - 1 = T_t \left( \hat{\beta}_{t-1}|t-2 + S_t^{-1}X_t' (Y_t - X_t \hat{\beta}_{t-1}|t-2) \right)
\]
where \( S_{t-1} = E[(\hat{\beta}_{t-1} - \hat{\beta}_{t-1})(\hat{\beta}_{t-1} - \hat{\beta}_{t-1})'] \). This estimator is sufficient from all the past data \( Y_1, Y_2, \ldots, Y_{t-1} \) in the sense of being a sufficient estimator, for the current state vector \( \hat{\beta}_t \).

Furthermore, the random vector \( (\hat{\beta}_t - \hat{\beta}_t|_{t-1}) \) is normally distributed with

\[
E(\hat{\beta}_t - \hat{\beta}_t|_{t-1}) = 0
\]

\[
E[(\hat{\beta}_t - \hat{\beta}_t|_{t-1})(\hat{\beta}_t - \hat{\beta}_t|_{t-1})'] = T_t S_{t-1} T_t' + Q_t.
\]

**Proof**

Let the probability density function of the random variable \( z = (A|B) \) be represented by \( f(A|B) \). Thus,

\[
f(\hat{\beta}_t|Y_{t-1}) = \int f(\hat{\beta}_t|\hat{\beta}_t|_{t-1}) f(\hat{\beta}_t|Y_{t-1}) d \hat{\beta}_{t-1}
\]

\[
= \int K_3 \exp\left\{-\frac{1}{2} P_t\right\} d \hat{\beta}_{t-1}.
\]

The individual conditional distributions are

\[
f(\hat{\beta}_t|\hat{\beta}_t|_{t-1}) = K_1 \exp\left\{-\frac{1}{2}(\hat{\beta}_t - T_t \hat{\beta}_{t-1})'Q_t^{-1}(\hat{\beta}_t - T_t \hat{\beta}_{t-1})\right\}
\]

and

\[
f(\hat{\beta}_t|Y_{t-1}) = K_2 \exp\left\{-\frac{1}{2}(\hat{\beta}_t - \hat{\beta}_t|_{t-1})'S_{t-1}^{-1}(\hat{\beta}_t - \hat{\beta}_t|_{t-1})\right\}
\]

where

\[
\hat{\beta}_{t-1} = T_{t-1} \hat{\beta}_{t-2} + S_{t-1} X_t' R_{t-1}^{-1} (Y_{t-1} - X_t T_t \hat{\beta}_{t-2})
\]

and

\[
S_{t-1} = [(T_{t-1} R_{t-1} T_t') + Q_{t-1}]^{-1} + X_t' R_{t-1}^{-1} X_{t-1}^{-1}.
\]
Therefore, the exponent $P_t$ is

$$P_t = (\beta_{t-1} - T_t \hat{\beta}_{t-1})'Q_t^{-1} (\beta_{t-1} - T_t \hat{\beta}_{t-1}) + (\beta_{t-1} - T_t \hat{\beta}_{t-1})'S_t^{-1} (\beta_{t-1} - T_t \hat{\beta}_{t-1}).$$

By completing the square, we can write

$$P_t = (\beta_{t-1} - T_t \hat{\beta}_{t-1} - T_t (\beta_{t-1} - T_t \hat{\beta}_{t-1}))'Q_t^{-1}$$

$$+ (\beta_{t-1} - T_t \hat{\beta}_{t-1})'S_t^{-1} (\beta_{t-1} - T_t \hat{\beta}_{t-1}) +$$

$$+ (\beta_{t-1} - T_t \hat{\beta}_{t-1})' (S_t^{-1} + T_t Q_t^{-1} T_t)(\beta_{t-1} - T_t \hat{\beta}_{t-1}) -$$

$$- 2(\beta_{t-1} - T_t \hat{\beta}_{t-1})'T_t Q_t^{-1} (\beta_{t-1} - T_t \hat{\beta}_{t-1}).$$

(A1.1) Let $M_t = (S_t^{-1} + T_t Q_t^{-1} T_t)^{-1}$. Then, we find

$$P_t = (\beta_{t-1} - T_t \hat{\beta}_{t-1})' (Q_t^{-1} Q_t^{-1} T_t M_t T_t Q_t^{-1}) (\beta_{t-1} - T_t \hat{\beta}_{t-1}) +$$

$$+ (\beta_{t-1} - M_t T_t Q_t^{-1} (\beta_{t-1} - T_t \hat{\beta}_{t-1}))'M_t^{-1}.$$

A theorem from matrix algebra shows us that we can write

$$Q_t^{-1} - Q_t^{-1} T_t M_t T_t Q_t^{-1} = (Q_t + T_t S_t^{-1} T_t')^{-1}$$

after recalling the equality from (A1.1) Then integrating with respect to $\beta_{t-1}$.
\[
\int K_3 \exp\left(-\frac{1}{2} P_t\right) d \beta_{t-1} = K_4 \exp\left(-\frac{1}{2} P^*_t\right)
\]

where \( P^*_t = (\beta_t - T^t \hat{\beta}_{t-1})'(Q_t + T^t T_{t-1})^{-1} (\beta_t - T^t \hat{\beta}_{t-1}) \).

Therefore, the distribution \( f(\beta_t | Y_{t-1}) \) is normal with mean

\[(A1.2) \quad E(\hat{\beta}_t | Y_{t-1}) = T^t \hat{\beta}_{t-1}\]

and covariance matrix \((Q_t + T^t T_{t-1} T^t)\). From (A1.2), \( \hat{\beta}_t | Y_{t-1} = T^t \hat{\beta}_{t-1} \) is a sufficient estimator from the data \( Y_1, Y_2, \ldots, Y_{t-1} \) for the state vector \( \beta_t \).
Appendix 2

Suppose we assume the recursive model from Chapter 2 section 1. That is, we shall let our prior data be represented by

\[ R^* = \beta + \eta \]

and the experimental data be represented by

\[ \mathbf{Y} = X\beta + \varepsilon. \]

Furthermore, we shall assume

\[ \eta \text{ is } N(0, \Gamma), \quad \varepsilon \text{ is } N(0, \frac{1}{\mathbf{P}}) \]

and that the vectors \( \eta \) and \( \varepsilon \) are uncorrelated. The state vector \( \beta \) will be considered a random vector in this case. It will then be shown that the mean of the posterior distribution \( E(\hat{\beta} | \mathbf{Y}) \) is the same as the weighted least squares estimator from (2.4).

Bayes' theorem expresses the posterior distribution in terms of the prior and experimental information as

\[ f(\beta | \mathbf{Y}) = \frac{f(\mathbf{Y} | \beta) f(\beta)}{f(\mathbf{Y})}. \]

The probability densities \( f(\mathbf{Y} | \beta) \) and \( f(\beta) \) are readily obtained from (A2.1) and (A2.2). However, because \( \beta \) is random with an unknown covariance matrix, the distribu-
tional form of $f(Y)$ is unknown. Thus, we will rewrite (A2.3) to read

$$f(\beta | Y) \cdot f(Y) = f(Y | \beta) \cdot f(\beta).$$

From the known form on the right hand side of (A2.4), we will find an expression whose quadratic form may be written

$$Q = Q_1 + Q_2$$

where $Q_1$ involves $\beta$ and $Y$ and $Q_2$ does not involve $\beta$. Then,

$$Q_1 = \text{quadratic form of } f(\beta | Y), \text{ and }$$

$$Q_2 = \text{quadratic form of } f(Y).$$

From (A2.1) and (A2.2), we find that

$$f(Y | \beta) = f(\epsilon) = f(Y - X\beta) =$$

$$= K_1 \exp\left\{- \frac{1}{2} (Y - X\beta)' \sum^{-1} (Y - X\beta) \right\}, \text{ and }$$

$$f(\beta) = K_2 \exp\left\{- \frac{1}{2} (\beta - \beta^*)' \Gamma^{-1} (\beta - \beta^*) \right\}.$$  

Substituting into (A2.4), we have

$$f(\beta | Y) f(Y) = K_1 K_2 \exp\left\{- \frac{1}{2} Q \right\}$$

where $Q = (Y - X\beta)' \sum^{-1} (Y - X\beta) + (\beta - \beta^*)' \Gamma^{-1} (\beta - \beta^*)$. 
Suppose we expand $Q$ and write

$$Q = \frac{1}{\Gamma^{-1}\beta} - \beta'x'\frac{1}{\Gamma^{-1}x\beta} - \frac{1}{\Gamma^{-1}x\beta} + \beta'x'\frac{1}{\Gamma^{-1}x\beta} \]

Now, let

$$Q = Q + (\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'\beta}(\Gamma^{-1} + x' \frac{1}{\Gamma^{-1}x'}\beta)^{-1}(\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'}\beta) - (\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'}\beta)'(\Gamma^{-1} + x' \frac{1}{\Gamma^{-1}x'}\beta)^{-1}(\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'}\beta) + x' \frac{1}{\Gamma^{-1}x'}\beta \gamma^{-1}\beta - (\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'}\beta)'(\Gamma^{-1} + x' \frac{1}{\Gamma^{-1}x'}\beta)^{-1}(\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'}\beta)) \]

Gathering terms, we have

$$Q = \{\beta' - (\Gamma^{-1} + x' \frac{1}{\Gamma^{-1}x'})^{-1}(\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'})\} + \{x' \frac{1}{\Gamma^{-1}x'} + \beta^*\gamma^{-1}\beta - (\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'})'(\Gamma^{-1} + x' \frac{1}{\Gamma^{-1}x'})^{-1}(\Gamma^{-1}\beta^* + x' \frac{1}{\Gamma^{-1}x'})\}. (A2.5) \]

Suppose we let $Q = Q_1 + Q_2$ where $Q_1$ and $Q_2$ are the respective terms in the brackets from (A2.5). Since $\beta$ appears only in $Q_1$, then
f(\mathbf{\bar{y}} | Y) = K_3 \exp\{-\frac{1}{2}Q_{\mathbf{1}}\}.

Thus, the random vector \( \mathbf{\bar{y}} | Y \) is normal with mean vector

\[
E(\mathbf{\bar{y}} | Y) = (\Gamma^{-1} + X' \mathbf{\Sigma}_Y^{-1}X)^{-1}(\Gamma^{-1} \mathbf{\bar{y}} + X' \mathbf{\Sigma}_Y^{-1}Y)
\]

and covariance matrix

\[
V(\mathbf{\bar{y}} | Y) = (\Gamma^{-1} + X' \mathbf{\Sigma}_Y^{-1}X)^{-1}
\]

provided that \( (\Gamma^{-1} + X' \mathbf{\Sigma}_Y^{-1}X)^{-1} \) is positive definite.

This discussion also applies to the Kalman model by performing the transformations from (2.2.5).
Appendix 3

In this section we examine the behavior of the parameter vector estimator under different modelling restrictions.

Lemma 1

In the recursive linear model, the added corrections to the parameter vector estimator due to the addition of new experimental data have expectation zero.

Proof

From Chapter 2, the weighted least squares estimator for the parameter vector $\beta$ is

$$\hat{\beta}(\Sigma, \Gamma) = \hat{\beta}^* + \psi X' \Sigma^{-1} (y - X\hat{\beta}^*)$$

where $\hat{\beta}^*$ is the estimator of the parameter vector obtained from prior experimental data. Thus, the addition due to the new experimental data is

$$\psi X' \Sigma^{-1} (y - X\hat{\beta}^*)$$

Taking the expectation of this term

$$E[\psi X' \Sigma^{-1} (y - X\hat{\beta}^*)] = \psi X' \Sigma^{-1} E(y - X\hat{\beta}^*)$$

$$= \psi X' \Sigma^{-1} E(X\hat{\beta} + \epsilon - X\hat{\beta}^*)$$

$$= \psi X' \Sigma^{-1} E(X\hat{\beta} + \epsilon - X(\hat{\beta} + \eta))$$

$$= \psi X' \Sigma^{-1} E(\epsilon - X\eta)$$

$$= 0.$$
Lemma 2
In the completely random recursive model (Kalman), the state vector estimator is not unbiased in the usual sense. In particular,

\[ E(\hat{\beta}_t) = E(\hat{\beta}_t) = \phi_{t,0}\beta_0 \]

where \( \phi_{t,0} = T_t \cdot T_{t-1} \cdots T_2 \cdot T_1 \).

Proof
Consider the Kalman model from Chapter 2, section 2. The least squares estimator for \( \beta_t \) is

\[ \hat{\beta}_t = \hat{\beta}_t|t-1 + M_t(Y_t - X_t\hat{\beta}_t|t-1), \]

where the smoothing matrix \( M_t \) is

\[ M_t = S_t X_t' R_t^{-1}. \]

Taking the expectation of both sides,

\[ E(\hat{\beta}_t) = E(\hat{\beta}_t|t-1) + M_t E(X_t\hat{\beta}_t + u_t - X_t\hat{\beta}_t|t-1) \]

\[ = E(\beta_t + \epsilon_t) + M_t E(X_t\beta_t - X_t(\beta_t + \epsilon_t)) \]

\[ = E(\beta_t) + 0 \]

\[ = E(\beta_t), \]

since Appendix 1 shows \( (\hat{\beta}_t|t-1 - \beta_t) \) is \( N(0, S_t|t-1) \). Since the state vector \( \beta_t \) is random and
\[ \beta_t = T_t \beta_{t-1} + v_t \]

then

\[ \beta_t = T_t T_{t-1} \cdots T_2 T_1 \beta_0 + \sum_{i=1}^{t} \phi_{t,i} v_i \]

where \( \phi_{t,t} = I \). Therefore,

\[
E(\hat{\beta}_t) = E[T_t T_{t-1} \cdots T_2 T_1 \beta_0 + \sum_{i=1}^{t} \phi_{t,i} v_i] = \phi_{t,0} \beta_0.
\]

**Lemma 3**

For the stochastic linear model with a nonrandom state vector,

\[
|\hat{\beta}_t - \beta_{t-1}| \xrightarrow{\text{Pr}} 0 \quad \text{as} \quad t^{3/2} \to \infty
\]

**Proof**

From the model referenced as (2.3.1), the weighted least squares estimator for \( \beta_t \) is

\[
\hat{\beta}_t = \beta_t |_{t-1} + S_t X'_t R_t^{-1}(Y_t - X_t \hat{\beta}_t |_{t-1}).
\]

Thus, the "added correction" at time \( t \) is

\[
\hat{\beta}_t - \beta_t |_{t-1} = S_t X'_t R_t^{-1}(Y_t - X_t \hat{\beta}_t |_{t-1}) = S_t X'_t R_t^{-1}(X_t (\beta_t - \hat{\beta}_t |_{t-1}) + u_t).
\]

The covariance matrix for this expression is
(A3.1) \( \text{cov}(\hat{\beta}_t - \hat{\beta}_{t-1} | t-1) = (S_t' R_t^{-1}) (X_t^t S_t | t-1 X'_t + R_t) (S_t' R_t^{-1})' \)

where \( S_t | t-1 = T_t S_{t-1} T'_t \).

The terms \((S_t X_t' R_t^{-1})\) are each of \( O(t^{-1}) \) as shown in section 2.3.1. The term \((X_t S_t | t-1 X'_t + R_t)\) is of \( O((t-1)^{-1}) \) since \( R_t \) and \( X_t \) are \( O(1) \) and \( S_t | t-1 = T_t S_{t-1} T'_t \) is of \( O((t-1)^{-1}) \).

If we then assume \( R_t \) is bounded, the covariance (A3.1) is of order \( O(t^{-3}(1+t)) \) or approximately of order \( O(t^{-3}) \). Applying the Cauchy-Schwartz inequality to the Euclidean norm of (A3.1), we find

\[
||\text{cov}(\hat{\beta}_t - \hat{\beta}_{t-1} | t-1)|| \leq ||S_t X_t R_t^{-1}|^2 ||X_t S_t | t-1 X'_t + R_t||
\]

and hence

(A3.2) \( ||\text{cov}(\hat{\beta}_t - \hat{\beta}_{t-1} | t-1)|| \to 0 \) as \( t^{3/2} \to \infty \)

By applying a multivariate form of the Chebycheff inequality (see Birnbaum and Marshall (3)) to A.3.2, we find that

\[
|\hat{\beta}_t - \hat{\beta}_{t-1} | t-1 \xrightarrow{Pr} 0, \text{ as } t^{3/2} \to \infty.
\]
Lemma 4

For the stochastic linear model with a nonrandom state vector,

\[ |\hat{\beta}_t - \beta_t| \xrightarrow{\text{Pr}} 0 \quad \text{as} \quad t^{1/2} \to \infty. \]

Proof

From (2.3.4), \( |S_t| \to \{0\} \), the null matrix, as \( t \to \infty \).

Applying the Chebycheff inequality we find

\[ |\hat{\beta}_t - \beta_t| \xrightarrow{\text{Pr}} 0 \quad \text{as} \quad t^{1/2} \to \infty. \]
Appendix 4

The essence of this section, due to Duncan (8), notes that the usual fixed state vector theory of weighted least squares can be applied to the Kalman model. The idea is an extension of the partitioned general linear model discussion from Chapter 2. Two modifications necessary to motivate this discussion are that

i) the prior means of the random state vector be treated as additional observations, and

ii) the variance of the estimators be taken as that of the deviations $\hat{\beta}_t - \beta_t$ from the true variable state being estimated and not that of $\hat{\beta}_t$ alone.

Let the model be represented by the recursive equations

$$ Y_t = X_t \beta_t + u_t, \quad \text{and} $$
$$ \beta_t = T_t \beta_{t-1} + v_t $$

where

$u_t$ is $N(0, R_t)$, $v_t$ is $N(0, Q_t)$

and where $Y_t$: (nx1), $\beta_t$: (px1). The vectors $u_t$ and $v_t$ will be assumed uncorrelated. The prior data is represented by its sufficient statistic $\hat{\beta}_t | t-1$, where

$$ \hat{\beta}_t | t-1 = T_t \beta_{t-1} = \beta_t + \xi_t $$

and $\xi_t$ is $N(0, S_t | t-1)$, $E(u_t \xi_t') = 0$. 

Then, we can combine this prior and current information into a single pseudo model in which $\beta_t$ may now be considered as fixed. The pseudo model is

$$Y_{pt} = X_{pt}\beta_{t} + \epsilon_{pt}$$

which, in its complete form, is

$$\begin{bmatrix}
Y_t \\
\hat{\beta}_t | t-1
\end{bmatrix} = \begin{bmatrix} X_t \\
I
\end{bmatrix}\beta_t + \begin{bmatrix} u_t \\
\xi_t
\end{bmatrix}.$$ 

The results may be stated in the form of a Theorem.

**Theorem**

Given that the previous assumptions apply,

1. the maximum likelihood estimator for $\beta_t$ under the Kalman model (2.2.1) is the familiar weighted least squares estimator

$$\hat{\beta}_t = (X'_{pt} \omega^{-1} X_{pt})^{-1} X'_{pt} \omega^{-1} Y_{pt}$$

obtained from the pseudo model (A4.1), and

2. The covariance matrix of $\hat{\beta}_t - \beta_t$ is the same under the Kalman model and the pseudo model. In particular,

$$\text{cov}(\hat{\beta}_t - \beta_t) = (X'_{pt} \omega^{-1} X_{pt})^{-1}$$

where the weight matrix is
\[ \omega_p = \begin{bmatrix} R_t & 0 \\ 0 & S_{t|t-1} \end{bmatrix}. \]

Proof

Parts (i) and (ii) of the theorem can be proven by showing that the quadratic forms in the exponent of the density functions for models (2.2.1) and (A4.1) are identical.

Suppose for the Kalman model

\[ f(y_t | \beta_t, \hat{\beta}_{t|t-1}) = K_1 \exp\left\{-\frac{1}{2}Q_K \right\} \]

and for the pseudo model

\[ f(y_{pt} | \beta_t) = K_2 \exp\left\{-\frac{1}{2}Q_p \right\}. \]

Then

\[ Q_K = (y_t - X_t \hat{\beta}_{t-t})' R_t^{-1} (y_t - X_t \hat{\beta}_t) \]

\[ + (\hat{\beta}_{t|t-1} - \beta_t)' S_{t|t-1}^{-1} (\hat{\beta}_{t|t-1} - \beta_t) \]

while for the pseudo model

\[ Q_p = (y_{pt} - X_{pt} \beta_t_p)' \omega_p^{-1} (y_{pt} - X_{pt} \beta_t) \]

\[ = \left( \begin{bmatrix} y_t \\ \hat{\beta}_{t|t-1} \end{bmatrix} - \begin{bmatrix} X_t \\ I \end{bmatrix} \beta_t \right)' \begin{bmatrix} R_t^{-1} & 0 \\ 0 & S_{t|t-1}^{-1} \end{bmatrix} \left( \begin{bmatrix} y_t \\ \hat{\beta}_{t|t-1} \end{bmatrix} - \begin{bmatrix} X_t \\ I \end{bmatrix} \beta_t \right). \]
\[ (Y_t - X_t \hat{\beta}_t)' R_t^{-1} (Y_t - X_t \hat{\beta}_t) \]
\[ + (\hat{\beta}_t | t-1 - \beta_t) S_t^{-1} (\hat{\beta}_t | t-1 - \beta_t). \]

Since \( Q_k = Q_p \), the Kalman estimator (2.2.4) may be found by the method suggested in (A4.2).

To prove part (ii), let \( V = (X_p t^{-1} X_p)^{-1} \). Then, we may write

\[ \hat{\beta}_t - \beta_t = V(X_p t^{-1} Y_t - t^{-1} \beta_t \]
\[ = V(X_p t^{-1} Y_t + S_t^{-1} \hat{\beta}_t | t-1 \beta_t | t-1 - V^{-1} \beta_t \]
\[ = V(X_p t^{-1} Y_t + S_t^{-1} \hat{\beta}_t | t-1 - V^{-1} \beta_t \]
\[ = V(X_p t^{-1} Y_t - S_t^{-1} \beta_t | t-1 (\beta_t - \hat{\beta}_t | t-1) \]
\[ \text{since } V^{-1} = X_p t^{-1} X_p = X_p t^{-1} X_p + S_t^{-1} t-1. \]

Thus,
\[ \hat{\beta}_t - \beta_t = V(X_p t^{-1} Y_t + S_t^{-1} t-1 \xi_t). \]

Therefore, its covariance matrix is
\[ \text{cov}(\hat{\beta}_t - \beta_t) = V[X_p t^{-1} R_t^{-1} X_t + S_t^{-1} t-1 S_t^{t-1} t-1] V \]
\[ = V(X_p t^{-1} X_t + S_t^{-1} t-1) V \]
\[ = V V^{-1} V \]
\[ = V. \]
\[ = \left( x_t' R_t^{-1} x_t + s_{t|t-1}^{-1} \right)^{-1} \]

\[ = \left( x_{pt}' \omega_p^{-1} x_{pt} \right)^{-1}. \]
BIBLIOGRAPHY


