COMPARISON OF FITTING TECHNIQUES FOR SYSTEMS OF FORESTRY EQUATIONS

By

Valerie LeMay

B. Sc. (Forestry), University of Alberta, 1981
M. Sc. (Forestry), University of Alberta, 1982

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF FORESTRY

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1988

© Valerie LeMay, 1988
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Forestry
The University of British Columbia
Vancouver, Canada

Date Dec 15, 1988
Abstract

In order to describe forestry problems, a system of equations is commonly used. The chosen system may be simultaneous, in that a variable which appears on the left hand side of an equation also appears on the right hand side of another equation in the system. Also, the error terms among equations of the system may be contemporaneously correlated, and error terms within individual equations may be non-iid in that they may be dependent (serially correlated) or not identically distributed (heteroskedastic) or both. Ideally, the fitting technique used to fit systems of equations should be simple; estimates of coefficients and their associated variances should be unbiased, or at least consistent, and efficient; small and large sample properties of the estimates should be known; and logical compatibility should be present in the fitted system.

The first objective of this research was to find a fitting technique from the literature which meets the desired criteria for simultaneous, contemporaneously correlated systems of equations, in which the error terms for individual equations are non-iid. This objective was not met in that no technique was found in the literature which satisfies the desired criteria for a system of equations with this error structure. However, information from the literature was used to derive a new fitting technique as part of this research project, and labelled multistage least squares (MSLS). The MSLS technique is an extension of three stage least squares from econometrics research, and can be used to find consistent and asymptotically efficient estimates of coefficients, and confidence limits can also be calculated for large sample sizes. For small sample sizes, an iterative routine labelled iterated multistage least squares (IMLS) was derived.

The second objective was to compare this technique to the commonly used techniques
of using ordinary least squares (simple or multiple linear regression and nonlinear least squares regression), and of substituting all of the equations into a composite model and using ordinary least squares to fit the composite model. The three techniques were applied to three forestry problems for which a system of equations is used. The criteria for comparing the results included comparing goodness-of-fit measures (Fit Index, Mean Absolute Deviation, Mean Deviation), comparing the traces of the estimated coefficient covariance matrices, and calculating a summed rank, based on the presence or absence of desired properties of the estimates.

The comparison indicated that OLS results in the best goodness-of-fit measures for all three forestry problems; however, estimates of coefficients are biased and inconsistent for simultaneous systems. Also, the estimated coefficient covariance matrix cannot be used to calculate confidence intervals for the true parameters, or to test hypothesis statements. Finally, compatibility among equations is not assured. The fit of the composite model was attractive for the systems tested; however, only one left hand side variable was estimated, and, for larger systems with more variables and more equations, this technique may not be appropriate. The MSLS technique resulted in goodness-of-fit measures which were close to the OLS goodness-of-fit measures. Of most importance, however, is that the MSLS fit ensures compatibility among equations, estimates of coefficients and their variances are consistent, estimates are asymptotically efficient, and confidence limits can be calculated for large sample sizes using the estimated variances and probabilities from the normal distribution. Also, the number and difficulty of steps required for the MSLS technique were similar to the OLS fit of individual equations. The main disadvantage to using the MSLS technique is that a large amount of computer memory is required; for some forestry problems with very large sample sizes, the use of a subsample or the exclusion of the final step of the MSLS fit were suggested. This would result in some loss of efficiency, but estimated coefficients and their variances would be consistent.
# Table of Contents

Abstract

List of Tables

List of Figures

Acknowledgement

1 Introduction

2 Previous Attempts at Fitting Forestry Equation Systems
   2.1 Ordinary Least Squares for Individual Equations
   2.2 Compatible Systems using Substitution
   2.3 Minimum Loss Function
   2.4 Econometric Methods Based on the Assumption of iid Error Terms
   2.5 Econometric Methods without the Assumption of iid Error Terms
   2.6 Discussion of Previous Approaches

3 An Alternative Simultaneous Fitting Procedure
   3.1 Extension of Econometric Methods to Multistage Least Squares
   3.2 Estimation of the Error Covariance Matrix
      3.2.1 Autocorrelation and Contemporaneous Correlation
      3.2.2 Heteroskedasticity and Contemporaneous Correlation
3.2.3 Autocorrelation, Heteroskedasticity, and Contemporaneous Correlation ................................................................. 50
3.3 Confidence Limits, Hypothesis Testing, and Prediction ................................................................. 54

4 Procedures for Comparison of Fitting Techniques ................................................................. 59
4.1 Selection of Equations for the Systems ................................................................. 59
4.2 Obtaining the OLS, Composite, and MSLS Fits ................................................................. 60
4.3 Tests for Heteroskedasticity and Serial Correlation ................................................................. 62
4.4 Criteria for Comparison of the Three Techniques ................................................................. 64
4.4.1 Goodness-of-fit Measures ................................................................. 64
4.4.2 Trace of the Estimated Coefficient Covariance Matrix ................................................................. 66
4.4.3 Table of Estimated Coefficients and Standard Deviations ................................................................. 67
4.4.4 Ranking for Other Features ................................................................. 67

5 Application 1: Tree Volume Estimation ................................................................. 69
5.1 Introduction ................................................................. 69
5.2 Preparation of Data ................................................................. 69
5.3 Model Selection ................................................................. 72
5.4 Ordinary Least Squares Fit ................................................................. 80
5.4.1 Unweighted Simple or Multiple Linear Regression ................................................................. 80
5.4.2 Testing for iid Error Terms ................................................................. 80
5.4.3 Estimating the Error Covariance Matrix of Each Equation ................................................................. 82
5.4.4 Appropriate OLS Fit Based on Error Structure ................................................................. 84
5.5 Composite Model Fit ................................................................. 84
5.5.1 Derivation of the Composite Model ................................................................. 84
5.5.2 Unweighted Regression of Composite Model ................................................................. 85
5.5.3 Testing for iid Error Terms ................................................................. 86
5.5.4 Estimating the Variance of the Error Terms .................................. 86
5.5.5 Weighted Regression of the Composite Model ................................. 87
5.6 MSLS Fit ......................................................................................... 88
  5.6.1 First Stage Equations .................................................................. 88
  5.6.2 Second Stage Equations ............................................................. 89
5.6.3 Testing for iid Error Terms .......................................................... 89
5.6.4 Estimation of the Error Covariance Matrix ..................................... 90
5.6.5 EGLS to Fit the System of Equations ......................................... 92
5.7 Comparison of the Three Fitting Techniques ..................................... 93
  5.7.1 Goodness-of-fit Measures .......................................................... 93
  5.7.2 Relative Variances ..................................................................... 97
  5.7.3 Table of Estimated Coefficients and Standard Deviations .......... 98
  5.7.4 Ranking for Other Features ...................................................... 100
5.8 Conclusion ..................................................................................... 102

6 Application 2: Estimation of Tree Diameter Distribution .................. 105
  6.1 Introduction ................................................................................. 105
  6.2 Preparation of Data ...................................................................... 107
  6.3 Model Selection .......................................................................... 110
  6.4 Ordinary Least Squares Fit .......................................................... 113
    6.4.1 Unweighted Simple or Multiple Linear Regression .................. 113
    6.4.2 Testing for iid Error Terms ................................................... 113
    6.4.3 Estimating the Error Covariance Matrix of Each Equation ....... 114
    6.4.4 Appropriate OLS Fit Based on Error Structure ..................... 116
  6.5 Composite Model Fit .................................................................... 116
    6.5.1 Derivation of the Composite Model ....................................... 116
6.5.2 Unweighted Regression of Composite Model 117
6.5.3 Testing for iid Error Terms and Weighted Regression 118

6.6 MSLS Fit 119
   6.6.1 First Stage Equations 119
   6.6.2 Second Stage Equations 119
   6.6.3 Testing for iid Error Terms 120
   6.6.4 Estimation of the Error Covariance Matrix 121
   6.6.5 EGLS to Fit the System of Equations 123

6.7 Comparison of the Three Fitting Techniques 124
   6.7.1 Goodness-of-fit Measures 124
   6.7.2 Relative Variances 129
   6.7.3 Table of Estimated Coefficients and Standard Deviations 130
   6.7.4 Ranking for Other Features 132

6.8 Conclusion 133

7 Application 3: Volume Growth and Yield 136
   7.1 Introduction 136
   7.2 Data Preparation 137
   7.3 Model Selection 138
   7.4 Ordinary Least Squares Fit 139
      7.4.1 Unweighted Simple or Multiple Linear Regression 139
      7.4.2 Testing for iid Error Terms 139
      7.4.3 Appropriate OLS Fit Based on Error Structure 140
   7.5 Composite Model Fit 140
      7.5.1 Derivation of the Composite Model 140
      7.5.2 Unweighted Regression of the Composite Model 141

vii
7.5.3 Testing for iid Error Terms ........................................ 141
7.5.4 Weighted Fit of the Composite Model ............................ 142

7.6 MSLS Fit ............................................................. 142
  7.6.1 First Stage Equations ............................................ 142
  7.6.2 Second Stage Equations ......................................... 143
  7.6.3 Testing for iid Error Terms .................................... 143
  7.6.4 Estimation of the Error Covariance Matrix ................... 144
  7.6.5 EGLS to Fit the System of Equations ......................... 144

7.7 Comparison of the Three Fitting Techniques ....................... 145
  7.7.1 Goodness-of-fit Measures ...................................... 145
  7.7.2 Relative Variances ............................................. 148
  7.7.3 Table of Estimated Coefficients and Standard Deviations ... 150
  7.7.4 Ranking for the Three Techniques ............................. 150

7.8 Conclusion .......................................................... 152

8 Overall Discussion and Conclusions .................................. 155

9 References Cited ..................................................... 160

A Glossary of Terms and Abbreviations ................................ 170

viii
List of Tables

2.1 Simultaneous Estimation Assuming iid Error Terms .......................... 21
5.2 Distribution of Selected Trees by Height and Dbh Classes .................... 71
5.3 Distribution of 100 Sample Trees ................................................... 73
5.4 Fit Indices for OLS and MSLS Fits of the Volume Equation System .......... 93
5.5 M.A.D. for Five Classes for the OLS Fit of the Volume Equation System 94
5.6 M.A.D. for Five Classes for the MSLS Fit of the Volume Equation System 95
5.7 M.A.D. for Merchantable Volume ..................................................... 95
5.8 M.D. for Five Classes for the OLS Fit of the Volume Equation System 96
5.9 M.D. for Five Classes for the MSLS Fit of the Volume Equation System 96
5.10 M.D. for Merchantable Volume ...................................................... 97
5.11 Trace of the Coefficient Covariance Matrix for Each Equation of the Volume Equation System ................................................................. 98
5.12 Estimated Coefficients and Standard Deviations of Coefficients for the Volume Equation System ............................................................. 99
5.13 Ranks for the Three Techniques ...................................................... 101
6.14 Distribution of Selected Psp's by Age and Stems per Hectare Classes ... 111
6.15 Fit Indices for OLS and MSLS Fits of the Dbh Distribution System ....... 124
6.16 M.A.D. for Five Classes for the OLS Fit of the Dbh Distribution System 126
6.17 M.A.D. for Five Classes for the MSLS Fit of the Dbh Distribution System 126
6.18 M.D. for Five Classes for the OLS Fit of the Dbh Distribution System 128
6.19 M.D. for Five Classes for the MSLS Fit of the Dbh Distribution System 128
6.20 Trace of the Coefficient Covariance Matrix for Each Equation of the Dbh Distribution System ................. 130
6.21 Estimated Coefficients and Standard Deviations of Coefficients for the Dbh Distribution System ................. 131
6.22 Ranks for the Three Techniques for the Fit of the Dbh Distribution System 132

7.23 Distribution of Selected Psps for the Third Application ................. 137
7.24 Fit Indices for OLS, Composite Model, MSLS and IMSLS Fits of the Yield Equation System ...................... 145
7.25 M.A.D. for Five Classes for the OLS Fit of the Yield Equation System 146
7.26 M.A.D. for Five Classes for the MSLS Fit of the Yield Equation System 146
7.27 M.A.D. for Five Classes for the Yield Composite Model Fit ...................... 147
7.28 M.D. for Five Classes for the OLS Fit of the Yield Equation System 147
7.29 M.D. for Five Classes for the MSLS Fit of the Yield Equation System 148
7.30 M.D. for Five Classes for the Yield Composite Model Fit ...................... 148
7.31 Trace of the Coefficient Covariance Matrix for Each Equation of the Yield Equation System ...................... 149
7.32 Estimated Coefficients and Standard Deviations of Coefficients for the Yield Equation System ...................... 150
7.33 Ranks for the Three Techniques ...................... 151
List of Figures

3.1 Extension of Econometric Least Squares Methods for Fitting Systems to MSLS ................................................................. 35
Acknowledgement

My greatest thanks go to the professors at University of British Columbia. In particular, Dr. A. Kozak, as well as the other members of my thesis committee, Drs. J.H.G. Smith, M. Bonner, P. Marshall, and D. Tait, of the Faculty of Forestry, and Dr. J. G. Cragg of the Department of Economics provided great assistance in completing my research. In addition, I would like to thank Dr. J. Wilson and Ms. N. Bertrams for their efforts in securing funding and for processing the necessary information.

Data were provided by the Timber Management Branch of the Alberta Forest Service. Special thanks go to Mr. D. J. Morgan, Forester-in-charge, Forest Measurement Section, and his staff members for their prompt response to my data needs and questions.

I would also like to thank Patrick Pelletier, my husband, for his patience and encouragement.
Chapter 1

Introduction

In order to describe forestry problems, a system of equations is frequently used. For instance, for the prediction of tree volume, tree height is first estimated, and then tree volume is estimated from measured diameter outside bark at breast height (dbh) and the estimated tree height. The desired parameter, tree volume, is therefore estimated through the use of a system of two equations.

The error structure for a system of equations will affect the results of any technique used to fit the system. The error terms among equations of the system may be correlated (contemporaneous correlation)\footnote{For definitions of terms and abbreviations used in this thesis, see Appendix A.} or variables appearing as dependent variables in one equation may appear as an independent variable in another equation of the system. For example, the system of equations for the prediction of tree volume may be chosen as follows:

\[
\begin{align*}
\text{height} &= \beta_0 + \beta_1 \text{dbh}^2 + \epsilon_1 \\
\log \text{volume} &= \log \beta_2 + \beta_3 \log \text{dbh} + \beta_4 \log \text{height} + \epsilon_2
\end{align*}
\]

where \textit{volume} is the volume of the main bole of the tree from ground to tree top;
\textit{height} is the height of the main stem from ground to tree top;
\textit{dbh} is the diameter outside bark measured at breast height, 1.3 metres above ground;
\textit{log} is the logarithm, base 10 or base e;
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ are coefficients to be estimated;  
$\epsilon_1, \epsilon_2$ are error terms.

Height is a stochastic variable which occurs as both a dependent variable on the left hand side (LHS) of an equation and also as an independent variable on the right hand side (RHS) of an equation. Also, a measure of tree taper such as form factor has been excluded from the system. Taper would likely affect both height and volume, hence both $\epsilon_1$ and $\epsilon_2$ include the error due to the exclusion of a measure of taper, and are therefore probably correlated (contemporaneous correlation). Further complications arise if the variances of the error terms for any of the equations in the system vary over the range of independent variables (heteroskedasticity\(^2\)), or if the error terms for a given equation are correlated with the previous error terms (serially correlation).

The error structure for a system of equations can therefore have any or all of the following characteristics.

1. Dependent variables may appear on the LHS of an equation in the system and also as a RHS variable in another equation so that the OLS assumption that the RHS variables are uncorrelated with the error term is not met for every equation. Systems with this characteristic are termed simultaneous equations.

2. The error terms among equations may be correlated indicating contemporaneous correlation.

3. Within individual equations, the error terms may be serially correlated (not independent), or the variances of the error terms may be heterogenous (not identically distributed), or both (neither independent nor identically distributed (non-iid)).

The fitting approach for systems of equations for forestry applications should ideally meet all of the following criteria.

\(^2\)This may also be spelled as *heteroscedasticity*. See McCulloch (1985) for discussion.
Chapter 1. Introduction

1. The routine should be simple in that few fitting steps are required.

2. Estimates of the coefficients and their associated variances should be unbiased or at least consistent. The estimates should have low variance (high efficiency).

3. Reported information on asymptotic and small sample properties of the estimated coefficients and their variances should be available.

4. Estimates should result in a compatible system of equations in that logical relationships among variables in the system should be maintained in the fitted system.

Criteria one through four should be met regardless of the error structure of the systems of equations (characteristics 1, 2, and 3, page 2).

The main objectives of this research were as follows:

1. To review forestry, econometrics, biometrics, and statistics literature and to choose, from this literature, a technique which satisfies all of the above criteria for fitting simultaneous, contemporaneously correlated systems of forestry equations, in which the error terms of individual equations are non-iid.

2. To compare this alternative technique to the most commonly used methods of (1) an appropriate OLS fit to each of the equations and (2) an appropriate OLS fit to a composite model created by substituting the equations of the system into one composite equation.

The central hypotheses of this thesis are: first, a fitting technique exists which satisfies the desired criteria for simultaneous, contemporaneously correlated systems of equations in which individual equations have non-iid error terms; and second, that any additional computational burden in using the technique is compensated by the benefits of meeting the desired criteria.
Chapter 1. Introduction

To meet the first objective, a literature search was conducted, first by examining forestry literature, and then by extending the search to econometrics, biometrics, and statistics literature. In order to restrict the scope of this thesis, only techniques based on least squares methodology were considered. The main alternative to the least squares approach is a maximum likelihood approach; however, the maximum likelihood approach requires that an assumption about the distribution of the dependent variables be made, and is more difficult to calculate. Also, the maximum likelihood approach is more sensitive to model specification error, to the presence of outliers, and to the presence of multicollinearity (Cragg, 1967; Summers, 1965). Least squares methodology was therefore selected as a more desirable method.

Objective one was not met, because the search of the literature failed to provide a fitting technique which satisfies all of the desirable properties for fitting systems with all three of the characteristics listed for the error structure. However, the information in the econometrics literature was used as the basis for the derivation of a new technique as part of this research. Results from several authors were combined into a comprehensive technique and labelled multistage least squares (MSLS) for this thesis. The MSLS technique is restricted to fitting systems of equations in which one sample set is used to fit the equations and, therefore, the number of samples is the same for every equation of the system. An iterated procedure was also derived and labelled iterated multistage least squares (IMLS).

To meet the second objective, the derived MSLS technique, and the two most commonly used techniques were used to fit a system of equations for each the following forestry problems.

---

3The use of the terms multistage least squares in this thesis refers only to the technique derived by this author. Pienaar and Shiver (1986) use the term multistage least squares in the summary of their paper; however, it is likely that they are referring to established econometric techniques for systems of equations where error terms for individual equations are iid, which can be considered to be a subset of the MSLS comprehensive technique described in this thesis.
Chapter 1. Introduction

1. The estimation of tree volume.

2. The estimation of diameter distributions from stand measurements, such as stand age and number of stems per unit area.

3. The estimation of volume and basal area growth and yield.

The systems of equations for these problems were expected to have different error structures. For the first problem, the error terms were expected to be heteroskedastic within equations and contemporaneously correlated among equations. The second problem involved the prediction of parameters of a probability distribution; each equation of the system estimated one of the parameters. For the third problem, the error terms were expected to be serially correlated and perhaps heteroskedastic within equations, and contemporaneously correlated among equations. These estimation problems are commonly encountered in forest management and endogenous variables tend to be estimates which are later expanded. For instance, the predicted volume on a given tree may be expanded to a per hectare estimate of volume.

The evaluation of the three fitting procedures, OLS, composite model, and MSLS, was based on comparing goodness-of fit measures, comparing the traces of the estimated coefficient covariance matrices, and calculating a summed rank, based on the presence or absence of desired properties of the estimates.

In order to distinguish between variable types, the terms used in econometrics are used. Stochastic variables which appear on the LHS of equations and may also appear on the RHS are termed endogenous variables derived from a Greek word meaning “generated from the inside” (Hu, 1973, page 121). Variables which appear only on the RHS are termed exogenous variables meaning “generated from the outside”. These exogenous variables were assumed to have very little error and could be considered to be fixed variables (nonstochastic).
Chapter 2

Previous Attempts at Fitting Forestry Equation Systems

2.1 Ordinary Least Squares for Individual Equations

For systems of equations in forestry, the most common fitting method has been the independent fitting of each equation in the system using an appropriate least squares procedure, such as simple linear regression, multiple linear regression, weighted regression, or nonlinear least squares (Burkhart, 1986; Furnival and Wilson, Jr., 1971). The OLS approach has appeal in that the method is well known and calculations to obtain estimates of coefficients and variances of coefficients and of dependent variables are relatively simple.

The standard OLS procedure of simple or multiple linear regression applied to individual equations yields estimates of coefficients which are best linear unbiased estimates (BLUE) for linear equations if the following assumptions are met.

1. The error terms for an equation are iid; serial correlation and heteroskedasticity are not present within individual equations.

2. The variables on the RHS of each equation are uncorrelated with the error term of the equation (nonstochastic).

3. The error terms among equations are not correlated, meaning that contemporaneous correlation is not present.
Chapter 2. Previous Attempts at Fitting Forestry Equation Systems

A system of three equations appears as follows:

\[
\begin{align*}
y_1 &= X_1 \beta_1 + \epsilon_1 \\
y_2 &= X_2 \beta_2 + \epsilon_2 \\
y_3 &= X_3 \beta_3 + \epsilon_3
\end{align*}
\]  
(2.3) (2.4) (2.5)

where \( y_i \) is an \( n \) by 1 matrix of the sample values for the \( i^{th} \) endogenous variable in the system of equations;

\( X_i \) is an \( n \) by \( k_i \) matrix of the sample values for all of the exogenous variables which affect the \( i^{th} \) endogenous variable in this equation of the system;

\( \beta_i \) is a \( k_i \) by 1 matrix of the true coefficients associated with the exogenous variables of this equation;

\( \epsilon_i \) is an \( n \) by 1 matrix of the error terms associated with each sample of the endogenous variable;

\( n \) is the number of samples.

These equations can also be expressed as follows:

\[
y = X\beta + \epsilon = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = X_1 \begin{bmatrix} 0_n \\ 0_n \\ 0_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} X_1 & 0_n & 0_n \\ 0_n & X_2 & 0_n \\ 0_n & 0_n & X_3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix} = XB + E
\]  
(2.6)

where \( 0_n \) is an \( n \) by \( n \) submatrix of zeros.

If all of these assumptions are met for the system of equations, the covariance matrix of the error terms of the system, is a diagonal matrix as follows:

\[
\Omega = \begin{bmatrix} \sigma_1^2 I_n & 0_n & 0_n \\ 0_n & \sigma_2^2 I_n & 0_n \\ 0_n & 0_n & \sigma_3^2 I_n \end{bmatrix}
\]  
(2.7)

where \( \Omega \) is the error covariance matrix for the system of equations;
\( \sigma_i^2 I_n \) is a diagonal submatrix representing iid error terms within the equation \( i \).

Each of the diagonal elements is equal to \( \sigma_i^2 \) for the \( i^{th} \) equation. \( I_n \) is the identity matrix of size \( n \) by \( n \);

\( 0_n \) is an \( n \) by \( n \) submatrix of zeros, indicating no contemporaneous correlation.

The covariance matrix then “falls apart” into separate matrices for each equation and simple or multiple linear regression can be used to fit each equation separately. The estimated coefficients will be BLUE, and also maximum likelihood estimates (MLE) if the error terms are normally distributed. The estimated coefficient covariance matrix also will be unbiased.

For individual equations, if the error terms are not identically distributed, in that their variances are not homogeneous, the error covariance matrix remains a diagonal matrix, but the diagonal terms for each equation are unequal, as shown below for three equations and three samples.

\[
\Omega = \begin{bmatrix}
\sigma_{11}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma_{12}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_{13}^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma_{21}^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_{22}^2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma_{23}^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma_{31}^2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{33}^2
\end{bmatrix}
\]  \( (2.8) \)

where \( \sigma_{ij}^2 \) is the variance for the \( i^{th} \) equation and the \( j^{th} \) sample.

Again, the covariance matrix can be divided into a separate matrix for each equation of the system. In this case, a specialized OLS technique such as a weighted regression
procedure or another transformation of the variables in each equation can be used so that the variance of the transformed regression is as the matrix in equation 2.7 (iid error terms), and a least squares procedure can be used to obtain estimates which are asymptotically BLUE for the untransformed variables (Judge et al., 1985, pages 420 and 421). Alternatively, the elements of the error covariance matrix can be estimated, and an estimated generalized least squares method (EGLS) could be used to fit each equation separately as follows:

\[ y_i = X_i \beta_i + \epsilon_i \]  \hspace{1cm} (2.9)

\[ \hat{\beta}_{i,\text{GLS}} = \left( X_i' \hat{\Phi}_i^{-1} X_i \right)^{-1} X_i' \hat{\Phi}_i^{-1} y_i \]  \hspace{1cm} (2.10)

where \( y_i \) is an \( n \) by 1 matrix of the sample values for the \( i^{th} \) endogenous variable in the system of equations;

\( X_i \) is an \( n \) by \( k_i \) matrix of the sample values for all of the exogenous variables which affect the \( i^{th} \) endogenous variable in this equation of the system;

\( \beta_i \) is a \( k_i \) by 1 matrix of the true coefficients associated with the exogenous variables of this equation;

\( \epsilon_i \) is an \( n \) by 1 matrix of the error terms associated with each sample of the endogenous variable;

\( \hat{\beta}_{i,\text{GLS}} \) is an estimate of the \( \beta_i \) matrix;

\( \hat{\Phi}_i \) is the estimated error covariance matrix for \( i^{th} \) equation of the system.

Generally, for EGLS, if the estimated error covariance matrix is consistent, the estimated coefficients and variances of these coefficients are consistent, and the distribution of each of the estimated coefficients is asymptotically normal. Several estimators for the error covariance matrix were presented by Judge and others (1985, pages 419 to 464) depending on the assumptions made concerning the heteroskedasticity. The procedure may be iterated by using the results of the EGLS fit to obtain a new estimate of the
error covariance matrix and from this obtain a new $\hat{\beta}_{GLS}$, and so on until the estimated coefficients converge. These iterations will result in what Malinvaud (1980, page 285) termed a *quasi-maximum likelihood* estimator of the coefficients and the covariance matrix. If normality of the error terms is assumed, then the estimators become MLE and the asymptotic covariance matrix will reach the Cramer-Rao lower bound for efficiency (minimum variance).

If serial correlation is present, the covariance matrix is no longer diagonal, because relationships within the sample appear. In this case, if the relationship can be defined such as a first order autoregressive process, in which the correlation between error terms declines geometrically as the time between disturbances increases, a term may be added to the regression equation or a transformation of the variables may be used (Cochrane and Orcutt, 1949; Kadiyala, 1968). This technique was used by Monserud (1984) for a single equation to describe height growth using stem analysis data. Alternatively, as with the case of heteroskedasticity, a consistent estimator of the error covariance matrix may be found, and used in an EGLS procedure (Judge et al., 1985, pages 283 to 286). Gregoire (1987) demonstrated this procedure for a single equation where permanent sample plots were correlated over time.

If endogenous variables appear on the RHS of one equation of the system and on the LHS of another equation of the system (simultaneous equations), the assumption that the variables on the RHS of each equation are not correlated with the error is not met. The independent OLS fit using simple or multiple linear regression will result in biased estimates of coefficients. To prove that this bias exists, equation 2.3 can be extended to include endogenous variables on the right hand side as follows:

$$y_i = Y_i\gamma_i + X_i\beta_i + \epsilon_i \quad (2.11)$$

where $y_i$ is an $n$ by 1 matrix of the sample values for the $i^{th}$ endogenous variable in the
system of equations;

\( Y_i \) is an \( n \) by \( g - 1 \) matrix of the sample values for all of the endogenous variables excluding the \( i_{th} \) endogenous variable;

\( \gamma_i \) is a \( g - 1 \) by 1 matrix of coefficients associated with the endogenous variables on the RHS of the equation. Coefficients are set to zero if the associated endogenous variables do not affect the \( i_{th} \) endogenous variable of the system;

\( X_i \) is an \( n \) by \( k \) matrix of the sample values for all of the exogenous variables of the system;

\( \beta_i \) is a \( k \) by 1 matrix of coefficients associated with the exogenous variables. Coefficients are set to zero if the associated exogenous variables do not affect the \( i_{th} \) endogenous variable of the system;

\( \epsilon_i \) is an \( n \) by 1 matrix of the errors associated with the \( i_{th} \) endogenous variable of the system.

The RHS variables and coefficients may be combined as follows:

\[
Y_i = [Y_i X_i] \begin{bmatrix} \gamma_i \\ \beta_i \end{bmatrix} + \epsilon_i = Z_i \delta_i + \epsilon_i
\tag{2.12}
\]

For the \( g \) equations of the system, using \( g = 3 \) equations,

\[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix} =
\begin{bmatrix}
  Z_1 & 0_n & 0_n \\
  0_n & Z_2 & 0_n \\
  0_n & 0_n & Z_3
\end{bmatrix}
\begin{bmatrix}
  \delta_1 \\
  \delta_2 \\
  \delta_3
\end{bmatrix}
+ \begin{bmatrix}
  \epsilon_1 \\
  \epsilon_2 \\
  \epsilon_3
\end{bmatrix}
\tag{2.13}
\]

which may be restated as follows:

\[
y = Z \Delta + E
\tag{2.14}
\]

where \( y \) is a \( gn \) by 1 matrix of \( n \) samples for each of \( g \) endogenous variables;

\( Z \) is a \( gn \) by \( g((g - 1) + k) \) matrix of the RHS variables;
\( \Delta \) is a \( g((g - 1) + k) \) by 1 matrix of the coefficients associated with the RHS variables;

\( E \) is a \( gn \) by 1 matrix of the residual error for \( n \) samples for each of \( g \) endogenous variables.

For a single equation, the OLS estimator of \( \delta_i \) is therefore:

\[
\hat{\delta}_i = (Z_i'Z_i)^{-1} Z_i'y_i
\]  

(2.15)

The expectation of the OLS estimator is:

\[
E[\hat{\delta}_i] = E\left[(Z_i'Z_i)^{-1} Z_i' (\delta_i + \epsilon_i)\right]
\]  

(2.16)

\[
E[\hat{\delta}_i] = \delta_i + E\left[(Z_i'Z_i)^{-1} Z_i' \epsilon_i\right]
\]  

(2.17)

The second term does not disappear as the \( Z_i \) are not independent of the error term and so the OLS fit is biased (Judge et al., 1985, page 571). This \textit{simultaneity bias} does not disappear if the sample size is increased. The estimates are also inconsistent (see proof in Judge et al., 1985, page 571). Also, since the estimated coefficients are biased and inconsistent, the estimated coefficient covariance matrix cannot be used to calculate confidence limits for the true coefficients. If confidence limits are calculated using these estimated variances of the coefficients, the limits would be incorrectly narrow.

If endogenous variables do not appear on the RHS, but contemporaneous correlation is present, the resulting OLS estimates of coefficients will not be most efficient as this correlation among equations is not included in the OLS fit of single equations. Zellner (1962) demonstrated this problem and called the system of equations with this contemporaneous correlation, seemingly unrelated regression (SUR) equations. If endogenous variables appear on the RHS and contemporaneous correlation is present, OLS estimates will again be biased and inconsistent, and, also, a loss in efficiency would result, because the information concerning correlation among equations would not be used in the OLS
fit. Finally, if contemporaneous correlation is present, endogenous variables appear on
the RHS, and the iid assumption is not met for one or more equations, the standard OLS
estimates using simple or multiple linear regression for individual equations will result
in biased and inconsistent estimates of the coefficients and their variances. In addition,
predictions of endogenous variables are not constrained to be logically compatible if an
OLS fit is used.

The practical implications of using the OLS fitting method for individual equations
are as follows:

1. For simultaneous systems, even though any bias between an estimated and the
   actual coefficient, or between a predicted and the actual value of the dependent
   variable may be small, these small biases may be magnified over a forest inventory.
   For instance, if volume per tree was in error by 0.10 cubic metres, and this tree
   represented 100 trees in the stand, the volume for the stand would be in error by
   10 cubic metres. Also, biases in certain coefficients may have a dramatic effect in
   the resulting estimates of dependent variables over the entire range or over partial
   ranges of the independent variables.

2. For simultaneous systems, estimates of coefficients based on the OLS fit of individ­
   ual equations are not only biased, they are inconsistent. Consequently, no matter
   how many samples are collected, there is no assurance that the sample estimates
   will be close to the population values for coefficients and dependent variables.

3. The estimated coefficient covariance matrix from the OLS fit of simultaneous equa­
   tions cannot be used to calculate confidence intervals. If these estimated variances
   are used, the confidence intervals will be incorrectly narrow, and there will be a
   higher chance that the true coefficients are not in the confidence interval. Also,
   hypotheses cannot be tested.
4. If contemporaneous correlation is present, OLS estimates will be less efficient, because the information concerning this correlation is not used in determining the OLS fit of individual equations. The resulting confidence limits will therefore be wider if OLS is used, and more samples will be required to obtain a desired precision.

5. Because the equations are fitted independently, compatibility of estimates is not assured. For instance, one equation may be used to estimate tree height and another equation may be used to estimate site index. If site index is defined as the height of the tree (or average tree) at 50 years measured at breast height, a desirable trait of this system of two equations would be that the height at 50 years from breast height predicted from the first equation, is equal to the site index of that tree as predicted from the second equation. The OLS fit of individual equations does not assure this logical compatibility.

The independent OLS approach has been termed the "naive" approach by Intriligator (1978, page 373), because information concerning the error structure of the system of equations is ignored. However, the method is still useful in that it is the easiest to calculate and computer programs are the most widely available. In addition, for preliminary work to define the system of equations, or where the system of equations is very large such as in systems for forestry growth and yield modelling, an independent OLS fit is probably the most practical method. However, compatibility is not assured, and estimates can be biased, inconsistent and not most efficient, depending on the error structure of the system.
2.2 Compatible Systems using Substitution

Another frequently used method for fitting systems of forestry equations is to use differentiation or substitution of variables and equations within the system. For instance, in an attempt to ensure logical compatibility of growth and yield estimates, Clutter (1963) suggested that growth equations should be obtained by taking the derivatives of the yield equations. Resulting estimates were not considered efficient. Bailey and Ware (1983) produced a compatible basal area growth and yield model by using the growth model to calculate yield. Matney and Sullivan (1982) described a system of substitutions to obtain compatible stock and stand tables. They estimated the parameters of the Weibull distribution by relating the integration of the distribution for volume and basal area to the predicted volume and basal area, thereby ensuring compatibility. Other applications of the method of substitution include Ramirez-Maldonado and others (1987), who developed a system of equations for predicting height growth and yield. A model to predict height at time one was first fitted, and then used to predict one of the coefficients of the model to estimate height at time two, so that estimates for growth or yield of height are compatible. McTague and Bailey (1987a) developed a compatible system of equations for basal area and diameter distribution by recovering the parameters of a Weibull distribution from predicted stand variables. First, the 10th, and 63rd percentiles, present and future, were predicted from site index, age and stems per hectare. The basal area was then predicted from these current percentiles. The "a" parameter of the Weibull distribution was predicted from age, number of stems, and the 10th percentile. The 90th percentile was then predicted from the 10th and 63rd percentiles, and site index and age. The "b" and "c" parameters were then calculated mathematically.

Substitution of all equations into one composite model has perhaps been even more widely used. Sullivan and Clutter (1972) developed a single linear model by substituting
the basal area growth equation into the volume yield equation to obtain a composite model. A maximum likelihood procedure was used to obtain unbiased estimates of the regression coefficients, because of serial correlation. The problem with this approach was that some of the variables from the original models were not significant in the composite model and disappeared from the resulting fit. The original biologically based models were lost; the result of changes in the variables was difficult to interpret. For example, the original model for predicting the log of volume at time two was as follows:

\[
\ln V_2 = f_1 (SI, 1/A_2, \ln BA_2)
\]

where \(\ln V_2\) is the natural logarithm of volume at time 2;
- \(SI\) is the site index;
- \(A_2\) is the age at time 2;
- \(BA_2\) is the basal area at time 2.

A substitution for basal area at time two resulted in the following equation.

\[
\ln V_2 = f_2 \left( SI, \frac{1}{A_2}, \frac{A_1}{A_2} \times \ln BA_1, 1 - \frac{A_1}{A_2}, \left(1 - \frac{A_1}{A_2}\right) \times SI \right)
\]

where \(A_1\) is the age at time 1;
- \(BA_1\) is the basal area at time 1.

The composite model no longer retained the expected biological relationship between volume and basal area. Also, even though the authors suggested that the combined linear model was compatible for prediction of growth and yield, the growth was obtained indirectly by subtracting yield at time two from yield at time one, rather than by an unique equation for growth and for yield.

A composite model for height growth based on the Chapman-Richard's model (Pienaar and Turnbull, 1973) is frequently proposed. The coefficients of the model are estimated for each tree or plot and then these coefficients are related to site index or
other sample observations such as habitat or region. The resulting equations to describe changes in the coefficients are then placed back into the Chapman-Richard’s equation to obtain a composite model (e.g. Beck, 1971; Graney and Burkhart, 1973; Lundgren and Dolid, 1970; Trousdell et al., 1974). Another example of the use of a composite model is VanDeusen and others (1982) who showed that a system of equations for merchantable volume to any height or to any diameter, total volume, and merchantable height could be made compatible by estimating one of the equations, and using this estimate for the other equations. Only one estimated coefficient was required. McTague and Bailey (1987b) proposed a more complex composite equation to estimate merchantable volume. The equation could be rearranged to estimate total volume, and for taper. The composite equation was fit, and the fitted coefficients were then used for each of the equations obtained by rearranging the composite model. For taper modelling, Kozak (in press) substituted an equation for the diameter at the inflection point, and another equation for the exponent of the equation, to obtain a composite model which described the diameter for a given height above ground.

The substitution method has been widely used to ensure compatibility of systems of forestry equations. However, the original biologically based models may be changed in the substitution, and neither efficiency nor unbiasedness of the estimated coefficients or their variances is assured. Also, only one variable on the LHS is estimated.

2.3 Minimum Loss Function

A simultaneous fit of two equations was obtained by Burkhart and Sprinz (1984) by minimizing a loss function which combined squared errors for the first equation with the squared errors for the second equation as follows:

\[ F = \sum \frac{(V_i - \hat{V}_i)^2}{\sigma_V^2} + \sum \frac{(B_i - \hat{B}_i)^2}{\sigma_B^2} \]  

(2.20)
where $V_i, \hat{V}_i$ is volume, actual and predicted;

$B_i, \hat{B}_i$ is basal area, actual and predicted;

$\hat{\sigma}_v^2$ is the mean square error from the OLS fit of the volume equation;

$\hat{\sigma}_b^2$ is the mean square error from the OLS fit of the basal area equation.

The advantage of this approach to fitting a system of equations is that constraints for coefficients across equations of the system can be introduced. Burkhart and Sprinz introduced the coefficients specified by Clutter (1963) to ensure compatibility between the predicted basal area and yield equations. Goodness-of-fit measures\(^1\) were used to test the resulting procedures. No estimate of small or large sample bias or efficiency was made, although Burkhart and Sprinz noted that the resulting estimates gave a higher sum of squared error for the first equation and a lower sum of squared error for the second equation than OLS. They also noted that the minimum loss function as defined above was lower with this simultaneous fit than with the OLS fit of the equations individually.

Reed and Green (1984) used a loss function similar to Burkhart and Sprinz (1984) to simultaneously estimate stem taper and volume coefficients, and constrained the coefficients so that logical relationships were represented by the coefficients. They tested four systems for estimating taper and volume, and found that the loss function decreased by 10 to 50 percent from the OLS fit of individual equations. Byrne and Reed (1986) extended the same loss function to a system of four equations for taper, total volume, volume ratio to an upper diameter, and volume ratio to an upper height.

Reed and others (1986) used a minimum loss function for fitting four basal area and yield equations based on Sullivan and Clutter (1972). Knoebel and others (1986) also fitted growth and yield equations using a minimum loss function to minimize the squared

---

\(^1\)The term *goodness-of-fit* measures is not to be confused with *goodness-of-fit* tests, such as the Chi-Square test for goodness-of-fit. The former refers to measures which indicate how well the estimated equation fits the sample data, whereas the latter refers to tests of whether the sample data are from a particular hypothesized distribution.
Chapter 2. Previous Attempts at Fitting Forestry Equation Systems

error for the system.

In all cases, the total sum of squares for the system represented by the loss function was reduced from the OLS fit of individual equations. Compatibility was introduced by constraining the coefficients, which may have resulted in biased estimates if these constraints are incorrect. If equations are simultaneous, these estimated coefficients will not only be biased, but also inconsistent as with the OLS fit of simultaneous equations. Also, the resulting estimates may not be most efficient.

2.4 Econometric Methods Based on the Assumption of iid Error Terms

The theory for simultaneously fitting a system of equations assuming iid error terms for individual equations was developed in the 1940s and early 1950s (Chow, 1983), and widely used in economics. The techniques have also been applied to agricultural problems (Friedman and Foote, 1955). Explanations of simpler techniques can be found in general econometric texts such as those by Gujarti (1978) and Wonnacott and Wonnacott (1979). More detailed explanations, proofs and other techniques can be found in Intriligator (1978) and Judge and others (1985). The assumption of iid error terms for individual equations results in a covariance matrix for the error terms of the system which are block diagonal as follows for a system of three equations.

\[
\Omega = \begin{bmatrix}
\sigma_{11}I_n & \sigma_{12}I_n & \sigma_{13}I_n \\
\sigma_{21}I_n & \sigma_{22}I_n & \sigma_{23}I_n \\
\sigma_{31}I_n & \sigma_{32}I_n & \sigma_{33}I_n
\end{bmatrix}
\] (2.21)

where \(\sigma_{ij}\) is the covariance of the error terms between equations if \(i \neq j\), or is the variance of the error terms if \(i = j\);

\(I_n\) is the identity matrix of size \(n\) by \(n\);

\(n\) is the number of samples.
When equations are independent, contemporaneous variances among equations are zero. The error covariance matrix becomes a diagonal matrix as in equation 2.7, because the $\sigma_{ij}$ terms which are not on the diagonal are zero. The simultaneous estimation of a system of equations reduces to OLS for each equation when the error covariance matrix is diagonal, and the equations are not simultaneous. A summary of common techniques for estimation of systems of equations based on least squares theory assuming iid error terms for individual equations is given in Table 2.1.

The condition specified for the 2SLS and 3SLS procedures is that the simultaneous system of equations must be *identified*. This condition is required so that an equation can be distinguished from other equations in the system. For just identified systems, if algebra is used to change the equations of the system (structural equations), so that the endogenous variables appear only on the LHS (reduced-form equations), coefficients estimated from an OLS fit of the reduced-form equations can be used to recover the coefficients of the structural equations. If the system is underidentified, there is not enough information available from the reduced-form coefficients to solve for the coefficients for the structural model. For overidentified systems, more than enough information is available from the reduced-form coefficients, so that more than one solution for the structural coefficients is possible.

If the simultaneous system of equations is underidentified, the statistical properties of 2SLS and 3SLS cannot be assumed. To test for identification, the rank condition must be met for each equation; an order condition can be used to define if the system is just-, over-, or underidentified, once the rank condition is met. First, a matrix of all of the coefficients of the system is constructed, with each column representing one variable of the system and each row representing one equation. To test whether the $i^{th}$ equation satisfies the rank condition, the row corresponding to the coefficients of that equation is deleted, and then the columns corresponding with nonzero coefficients for that equation
<table>
<thead>
<tr>
<th>Error Structure</th>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. All assumptions of OLS are met. The error covariance matrix is diagonal (equation 2.7).</td>
<td>OLS</td>
<td>OLS applied to each equation.</td>
</tr>
<tr>
<td>2. Contemporaneous correlation due to common absence of important exogenous variables in two or more equations of the system. All RHS variables are exogenous.</td>
<td>Seemingly Unrelated Regressions (SUR); Linear or Nonlinear</td>
<td>Estimate the error matrix and use EGLS or nonlinear equivalent, extended to a system of equations.</td>
</tr>
<tr>
<td>3. Endogenous variables on LHS and on RHS of equations in the system (Simultaneous system of equations).</td>
<td>Two stage Least Squares, Linear (2SLS), or Nonlinear (2SNLS)</td>
<td>Stage one: Regress each endogenous variable on all exogenous variables of the system. Stage two: Regress original system equations but replace endogenous variables on RHS with predicted endogenous variables from Stage one.</td>
</tr>
<tr>
<td>4. As type 3 above but contemporaneous correlation due to common absence of important exogenous variables from two or more equations of the system.</td>
<td>Three stage least squares, linear (3SLS), or nonlinear (3SNLS)</td>
<td>Stages one and two as 2SLS above. Stage three: Use the estimated error terms from Stage two to obtain an estimate of the error covariance matrix. EGLS for the system.</td>
</tr>
</tbody>
</table>
are deleted. The determinate of at least one submatrix of order \( g - 1 \) from this reduced matrix must be nonzero for the rank condition to be satisfied. The order condition can then be checked. If the number of exogenous variables in the system minus the number of exogenous variables in the \( i^{th} \) equation, is equal to the number of endogenous variables in the \( i^{th} \) equation minus one, the equation is justidentified. If this difference is greater, the equation is overidentified. All of the equations of the system must be identified for the system to be identified (See Gujarati, 1978, pages 353 to 365, for more explanation).

The earliest forestry application of econometric techniques for fitting systems of equations is credited to Furnival and Wilson (1971). They showed that the simultaneous fitting methods developed for economics could be applied to a system of equations to estimate growth and separately to a system of equations to estimate yield. They indicated that the advantage of the use of these methods over substitution or the OLS fit of individual equations is that confidence limits for coefficients of the system could be estimated. They noted that three stage least squares\(^2\) (3SLS) gave unexpectedly small elements in the estimated coefficient covariance matrix and so confidence limits would be narrow. Subsequently, the application of econometric theory for simultaneous estimation of coefficients in forest growth and yield equations, assuming that individual equations have iid error terms, has been demonstrated by several authors including Murphy and Sternitzke (1979), Murphy and Beltz (1981), Murphy (1983), Borders and Bailey (1986), and Hans (1986).

Murphy and Sternitzke (1979) and Murphy and Beltz (1981) used 3SLS to estimate the coefficients of growth and yield models for pine in the West Gulf Region. The selected equations were those developed by Clutter (1963) and Sullivan and Clutter (1972). The study by Furnival and Wilson (1971) was cited by them and used as the main reason for

\(^2\)The terms two stage or three stage least squares are not to be confused with the terms two or three stage sampling. The former refers to techniques used to analyze data to estimate coefficients whereas the latter terms refer to sampling designs.
opting to use a simultaneous approach to estimation of coefficients rather than an OLS approach to individual equations. Murphy and Sternitzke stated that no attempt was made to compare the simultaneous approach to the OLS approach. In both papers, the fit statistics and a discussion on how to use the resulting system were given. No tests were done to determine if individual equations were iid even though this is an assumption of the 3SLS procedure.

Murphy (1983) chose a system of nonlinear equations to model growth and yield. For each of the dependent variables in the system, he used nonlinear least squares to select the "best" model. He expected both contemporaneous and serial correlation in the system and so he chose the seemingly unrelated nonlinear regressions technique to estimate the coefficients of the chosen system of equations, even though SUR techniques assume that individual equations have iid error terms and are therefore not serially correlated. He then compared the simultaneous fit to the independent fitting of each equation in the system using nonlinear least squares. He concluded that the simultaneous fit had no evident benefit over the individual nonlinear least squares fit for the data and models tested.

Borders and Bailey (1986) noted that the advantages of the simultaneous fitting methods were that:

1. Point estimates are consistent and can be also efficient.

2. Compatibility of equations is obtained.

3. Interval estimates can be derived.

The work of Clutter (1963) was used as the basis for the establishment of a system of growth and yield equations. They then compared the OLS fit of each of the equations independently to the simultaneous fit of the system of equations. Three simultaneous
fitting procedures were used. First, a 2SLS procedure was used, followed by a 3SLS procedure, and then a 3SLS procedure, modified by restricting some of the coefficients, was used. They also estimated the covariance matrices of the estimated coefficients and of the predicted endogenous variables. They concluded that the estimates of the coefficients were much the same for the OLS, 2SLS, 3SLS, and restricted 3SLS techniques; however, the 3SLS fits were more efficient (lower variance). Also, compatibility was obtained with the 2SLS, 3SLS, and restricted 3SLS techniques.

Hans (1986) compared the OLS fit to a substitution method, and a SUR fit for Clutter's (1963) yield equation system. Goodness-of-fit measures for the original and an independent data set, were used to compare the alternative methods. SUR gave lower standard errors than OLS, but the goodness-of-fit measures were better for the OLS fit. Hans recognized that an alternative simultaneous fitting technique would have been more appropriate, because endogenous variables in the system also appear as RHS variables.

Amateis and others (1984) demonstrated that forestry applications of simultaneous fitting techniques other than for growth and yield exist. They compared the OLS fit to two different simultaneous fitting techniques for estimates of product yield. Three systems were established, based on differing assumptions of inter-relationships of product yield. The first system assumed no inter-relationships among equations (contemporaneous variances were assumed to be zero) and so an OLS fit was appropriate. The second system assumed that relationships proceeded in succession, meaning that the results from the first equation affected the second equation, and the results from the first and second equations affected the third equation and so on. The resulting system was therefore identified as a recursive system. Errors between successive equations were also assumed to be independent and so an OLS fit was used to fit each equation of the recursive system. The third system was a simultaneous system of equations with endogenous variables appearing on both the RHS and the LHS and errors between equations were assumed to be
correlated. This system was fitted using 3SLS. Goodness-of-fit measures were calculated for each fitted system and compared. A summed rank was calculated by comparing the predicted value from each fitted system, to the observed value. The fitted system was given a rank of one if the predicted value was closest to the observed value; other fitted systems were given a rank of zero. The ranks were summed over all observations; the 3SLS had a higher summed rank. The recursive system which was fitted was considered to be inappropriate as dependencies between different products and different equations appeared to be significant. They concluded by saying that the simultaneous fitting approach did not substantially alter the estimates of coefficients, but the methods were more appealing.

Borders and others (1987b) developed a system of 12 equations to predict 12 percentiles of a diameter distribution from other percentiles and stand attributes. The system was fitted using SUR, as error terms among equations appeared to be contemporaneously correlated. The system was set up in sequence and so endogenous variables appearing on the RHS of the equation were predicted by a previous equation, justifying the use of the SUR method.

The SUR method was also used by Bailey and da Silva (1987) to fit basal area at time two from basal area at time one. Since overlapping measurement intervals were used to obtain coefficients for the model and contemporaneous correlation was expected to be significant, the SUR technique was used to fit coefficients for a system of equations. Each equation represented one measurement length.

2.5 Econometric Methods without the Assumption of iid Error Terms

If the assumption that the error terms of individual equations are iid does not hold, the error structure is no longer block diagonal as in equation 2.21, and techniques which
assume that the error is block diagonal will no longer be asymptotically most efficient, because information concerning the variances not on the diagonal is not included in the resulting fit. Techniques developed for fitting systems of equations in this case, still require that assumptions about the error covariance matrix be made, because not all elements of the $g_n$ by $g_n$ error matrix of the system of equations can be consistently estimated with only $n$ samples (Judge et al., 1985, page 174).

Few attempts have been made to obtain an efficient solution for forestry systems which have non-iid error terms for individual equations. The first attempt was by Ferguson and Leech (1978) to predict coefficients of a yield equation from exogenous variables outside of the system, a technique often labelled as parameter prediction in forestry literature. They noted that much statistical literature was available for fitting equations in which the coefficients were considered to be random variables, but little work had been done to fit equations in which the coefficients were considered to be random functions of other exogenous variables. To begin, they fitted a yield equation to each of 20 permanent sample plots, with nine observations each, using a standard OLS technique, multiple linear regression, resulting in a set of coefficients for each plot. The Durbin-Watson bounds test (Durbin and Watson, 1951) was applied to each plot to test for serial correlation. No significant correlations were noted on most plots; however, they did note that the tables published by Durbin and Watson showing the distribution of the test statistic did not include values for the small sample size of nine observations. In addition, variances of the error terms within each plot were considered to be homogenous, and so the mean square error from the standard OLS fit was used in estimating the variance of each coefficient on each plot. These estimated coefficient variances were examined and, using Bartlett’s test for equality of variances, it was determined that the variances differed significantly among plots. A system of three equations was established to predict each coefficient using stand variables. Since the three coefficients were related, the error terms
Chapter 2. Previous Attempts at Fitting Forestry Equation Systems

for these three equations were considered to be contemporaneously correlated. Multiple linear regression was then used to fit each of the random coefficient functions. An estimate of the error covariance matrix for the system of parameter prediction functions was found by combining the estimates of contemporaneous variance from this second OLS fit and the estimates of variance for each plot found by the OLS fit to the observations for each plot. The SUR approach was therefore modified by Ferguson and Leech to heteroskedastic, contemporaneously correlated equations. The error covariance matrix was assumed to have the following form, shown for three equations and three samples.

\[
\Omega = \begin{bmatrix}
\sigma_{111} & 0 & 0 & \sigma_{121} & 0 & 0 & \sigma_{131} & 0 & 0 \\
0 & \sigma_{112} & 0 & 0 & \sigma_{122} & 0 & 0 & \sigma_{132} & 0 \\
0 & 0 & \sigma_{113} & 0 & 0 & \sigma_{123} & 0 & 0 & \sigma_{133} \\
\sigma_{211} & 0 & 0 & \sigma_{221} & 0 & 0 & \sigma_{231} & 0 & 0 \\
0 & \sigma_{212} & 0 & 0 & \sigma_{222} & 0 & 0 & \sigma_{232} & 0 \\
0 & 0 & \sigma_{213} & 0 & 0 & \sigma_{223} & 0 & 0 & \sigma_{233} \\
\sigma_{311} & 0 & 0 & \sigma_{321} & 0 & 0 & \sigma_{331} & 0 & 0 \\
0 & \sigma_{312} & 0 & 0 & \sigma_{322} & 0 & 0 & \sigma_{332} & 0 \\
0 & 0 & \sigma_{313} & 0 & 0 & \sigma_{323} & 0 & 0 & \sigma_{333}
\end{bmatrix}
\]

(2.22)

where \(\sigma_{ijm}\) is the covariance for the \(i^{th}\) and the \(j^{th}\) equations and the \(m^{th}\) sample, estimated by \(\hat{\sigma}_{ijm}\);

the matrix is symmetric.

For the Ferguson and Leech study, the form of the matrix was expanded to include estimated variances for 20 plots. This estimated error covariance matrix was then used to obtain an EGLS fit, extrapolated to the whole system of equations as follows:³

\[
y = XB + E
\]

(2.23)

³EGLS extended to a system of equations has also been termed joint generalized least squares.
\[
\hat{\mathbf{B}}_{EGLS} = \left( \mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{y}
\]

(2.24)

\[
\text{Var}(\hat{\mathbf{B}}_{EGLS}) = \left( \mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right)^{-1}
\]

(2.25)

where \( \mathbf{y} \) is a \( gn \) by 1 matrix of the endogenous variables;

\( \mathbf{X} \) is a \( gn \) by \( gk \) matrix of the exogenous variables;

\( \mathbf{B} \) is a \( gk \) by 1 matrix of the coefficients associated with the exogenous variables.

Coefficients were set to zero if the associated exogenous variables did not affect the \( ih \) endogenous variable;

\( \mathbf{E} \) is a \( gn \) by 1 matrix of the error terms associated with \( n \) samples of each of \( g \) endogenous variables;

\( \hat{\mathbf{B}}_{EGLS} \) is an estimate of the \( \mathbf{B} \) matrix;

\( \hat{\mathbf{\Omega}} \) is a \( gn \) by \( gn \) matrix of the estimated error covariance matrix.

This overall fit accounted for heterogeneity of error within equations, as well as the correlation of error among equations by combining the error from the first and second step. All RHS variables were assumed to be nonstochastic, and serial correlation was assumed to be absent. Davis and West (1981) and Ferguson and Leech (1981) published notes of correction for the procedure given in Ferguson and Leech (1978).

Newberry (1984) and Newberry and Burkhart (1986) used the technique of Ferguson and Leech (1978), corrected by Davis and West (1981), to obtain estimates of coefficients of a taper function, but they expanded the technique to a nonlinear fitting method for the first step. A taper function was fitted for each tree using nonlinear least squares and variances for each coefficient were estimated using a Taylor’s series and a jackknife approximation. A runs test was then used to test for serial correlation (Lehman, 1975) for each tree. A sign test (Lehman, 1975) was then performed using the results of the runs tests, by counting the number of trees which showed serial correlation (significant results from the runs test), and then by comparing this count to the number of trees expected to
show serial correlation if the results were random. They decided that serial correlation was significant but stated that they were unaware of any technique to account for this serial correlation and so it was ignored. They then continued by relating each of the three coefficients to tree and stand characteristics to obtain a system of linear equations. The error covariance matrix accounting for the first (nonlinear) and second step (OLS on each of the parameter prediction equations) was estimated using the corrected Ferguson and Leech method, and EGLS was used to obtain estimates of coefficients and their associated variances for the system of equations. Some of the second step variance estimates were negative and so were given a value of zero. The authors noted no real improvements using the EGLS approach, extended to a system, to estimate coefficients for the parameter prediction system versus using OLS for each equation separately.

The Ferguson and Leech (1978) approach to parameter prediction, followed subsequently by Newberry (1984) and Newberry and Burkhart (1986), introduced a method for fitting a system of equations which accounted for heteroskedasticity (not identically distributed) and contemporaneous correlation. However, all RHS variables were exogenous, and serial correlation was assumed to be absent. Also, the LHS variables of the system were parameters obtained by fitting an equation for each sample unit (plot or tree). The variance of each of these LHS variables was therefore obtained from the initial estimation of the parameters. This method of determining the variance of each sample unit could not be applied to another system of equations in which the LHS variables are measured observations of endogenous variables, rather than estimated parameters.

2.6 Discussion of Previous Approaches

None of the methods used previously to fit systems of forestry equations has all of the desirable properties for fitting systems of equations identified in the introduction. Each
of the procedures lacks one or more of these properties, for a simultaneous, contemporaneously correlated system of equations, in which error terms of individual equations are non-iid.

The OLS fit of individual equations results in estimated coefficients which are biased, and inconsistent if the equations are simultaneous. Confidence limits for the true coefficients cannot be calculated using the estimated coefficient covariance matrix. For nonsimultaneous systems with non-iid error terms, weighted regression or EGLS can be used on each equation; however, if error terms are both non-iid and contemporaneously correlated, the resulting fit will not be most efficient. For simultaneous equations with non-iid error terms, the OLS fit or EGLS fit of individual equations results in biased and inconsistent estimates of the coefficients. Information on small sample properties in the presence of non-iid errors for simultaneous systems is limited, but information which is available for iid error terms for simultaneous systems shows that OLS is more biased for small samples than econometric methods, such as 2SLS and 3SLS, if equations are simultaneous (Cragg, 1967; Mikhail, 1975; Nagar, 1960; Sawa, 1969; Summers, 1965). Methods to increase the efficiency over the OLS fit, when contemporaneous correlation is significant are well documented (Kmenta and Gilbert, 1968; Zellner, 1962; Zellner and Huang, 1962) even for small sample sizes (Mehta and Swamy, 1976; Revankar, 1974; Zellner, 1963). In the presence of heterogeneity and serial correlation, the OLS fit of individual equations is less efficient relative to the generalized least squares (GLS) fit, and has also been shown to be less efficient relative to the estimated generalized least squares (EGLS) fit using consistent estimators of the error covariance matrix (Rao and Griliches, 1969). The OLS fit is therefore the simplest approach, thereby satisfying the first criterion, but the estimates can be biased, inconsistent, and not most efficient. In addition, the desired property of compatibility is not assured with the OLS fit.

The substitution method has an advantage over the OLS fit in that compatibility of
the system is achieved. However, no further advantages are gained through this method, unless the derived composite model has only exogenous variables on the RHS.

The minimum loss function approach has appeal in that compatibility is achieved, and, in addition, the mean square error for the system of equations may be reduced over the OLS fit individual equations, if biased estimates are introduced. The method is also relatively simple for small systems, but for large systems of many equations and many variables, the search to find the minimum value for the function may be long and difficult. Also, if the error terms are non-iid and contemporaneously correlated, the method is less efficient than alternative methods which make use of this information. If simultaneous equations are represented in the system, the resulting estimators will remain biased as with the OLS fit.

Econometric methods for fitting systems of equations based on the assumption that the error terms for individual equations are iid have the advantages over OLS of being compatible, consistent and less biased than OLS in the presence of simultaneity, and more efficient than OLS if contemporaneous correlation is significant. If the error terms of individual equations are non-iid, the methods result in inconsistent estimates of the coefficient covariance matrix, so that other techniques become more desirable.

The non-iid approach illustrated by Ferguson and Leech (1978) and Newberry and Burkhart (1986) for systems of parameter prediction equations are applications of EGLS to systems of equations. Ferguson and Leech (1978) indicated that their estimator for the error covariance matrix is consistent, and so the properties of GLS can be assumed for large sample sizes. This non-iid approach, therefore, yields consistent estimates of coefficients which are also asymptotically efficient (Judge et al., 1985, pages 175 and 176). The simultaneous fitting of the equations does ensure compatibility and the number of steps required is few. However, the technique to estimate the error covariance matrix
used by Ferguson and Leech is only applicable to a system of parameter prediction equations; error terms for an individual parameter prediction equation must not be serially correlated, and all RHS variables must be exogenous. Also, for small sample sizes such as that used by Ferguson and Leech, the level of efficiency is uncertain and variances may even exceed those of the OLS applied to each of the parameter prediction equations separately.
Chapter 3

An Alternative Simultaneous Fitting Procedure

3.1 Extension of Econometric Methods to Multistage Least Squares

A fitting technique which satisfies all of the desired criteria for simultaneous, contemporaneously correlated systems of equations with non-iid error terms, was not found in forestry literature, nor in the subsequent search of econometrics, biometrics, and statistics literature. Techniques for systems which have non-iid error terms for individual equations, and are contemporaneously correlated, but not simultaneous, were found for some forms of the non-iid error structure. For instance, the non-iid approach illustrated by Ferguson and Leech (1978) is an extension of the SUR technique for non-iid error terms, but this technique is only useful for a system of parameter prediction equations assuming heteroskedastic, contemporaneously correlated error terms. In econometric literature, techniques for extension of the SUR method to systems with serial correlation and contemporaneous correlation were found (Kmenta and Gilbert, 1970; Parks, 1967). However, extensions to heteroskedastic, contemporaneously correlated systems of equations which are not systems of parameter prediction equations were not found. Also, a technique for simultaneous systems in which error terms of individual equations are non-iid was not found.

Existing techniques were therefore extended to different error structures and to simultaneous systems as part of this research. This extended technique was labelled multistage least squares (MSLS), because the technique is based on least squares methodology, and
many fitting steps (stages) are required to obtain estimated coefficients and the associated covariance matrix.

The MSLS technique can be viewed as an extension of SUR techniques for non-iid error terms to simultaneous systems, and to other types of non-iid error structures not previously examined for SUR equations. It can also be viewed as an extension of the 3SLS technique to systems of simultaneous equations with non-iid error terms for individual equations. In order to show how the MSLS technique can be derived and justified from existing techniques, econometric methods assuming iid or non-iid error terms were combined into a flowchart presented in Figure 3.1, and described below.

The EGLS fitting technique was derived by Aitken (1934-35). This technique was extended to a system of equations with iid error terms for individual equations, for which contemporaneous correlation was significant by Zellner (1962) and labelled the SUR fitting method. Independently, Theil in 1953 (as referenced in Basmann, 1957) and Basmann (1957) derived a means of "purging" simultaneity bias from simultaneous equation systems called 2SLS. Zellner and Theil (1962) combined SUR and 2SLS for simultaneous, contemporaneously correlated systems, and called the combined technique 3SLS. Because 3SLS is an extension of the SUR technique to simultaneous equations, the error terms for individual equations are assumed to be iid. The steps for the 3SLS technique are as follows: ¹

Stage 1. Simple or multiple linear regression is used to predict endogenous variables from all of the exogenous variables in the system (first stage equations). Standard OLS techniques provide unbiased estimates of the coefficients to predict each of the endogenous variables even if the equations to predict these variables show heteroskedastic or serially correlated error terms.

¹These steps are also given in Table 2.1, but are included here so that the extension to equations with non-iid error terms can be shown.
Chapter 3. An Alternative Simultaneous Fitting Procedure

Figure 3.1: Extension of Econometric Least Squares Methods for Fitting Systems to MSLS

- Prediction using EGLS (Goldberger, 1962) (Yamamoto, 1979)
- EGLS for one equation (Aitken, 1934-35)
- SUR with Heteroskedasticity and Autocorrelation modified Parks (1967)
- Heteroskedasticity, SUR (Ferguson and Leech, 1978) modified Parks (1967) (Duncan, 1983)
- Autocorrelation, SUR (Parks, 1967) (Kmenta and Gilbert, 1970)
- 2SLS Theil in 1953 (Basmann, 1957)
- 3SLS (Zellner and Theil, 1962)
- Iterations (Dhrymes, 1971) (Magnus, 1978)
- Iterated Multistage Least Squares IMSLS
- Multistage Least Squares MSLS
Stage 2. To remove simultaneity bias, the predicted endogenous variables from stage 1 are substituted for the endogenous variables which appear on the RHS (second stage equations), and a standard OLS technique is applied to each of the modified equations of the system.\(^2\)

\[
y_i = [\hat{Y}_i X_i] \begin{bmatrix} \gamma_i \\ \beta_i \end{bmatrix} + \epsilon_i = \hat{Z}_i \hat{\delta}_i + \epsilon_i \tag{3.26}
\]

\[
\hat{\delta}_i = (\hat{Z}_i' \hat{Z}_i)^{-1} \hat{Z}_i' y_i \tag{3.27}
\]

This substitution ensures that all RHS variables are now independent of the error terms of each second stage equation so that the OLS fit for each equation will result in consistent estimates of all coefficients. The 2SLS technique is complete at this point.

Stage 3. To gain efficiency, for the 3SLS technique, the residual errors from the OLS fit of the second stage equations are used to obtain estimates of the contemporaneous variances using the estimator by Zellner (1962), as follows:

\[
\hat{\sigma}_{ij} = \frac{\hat{\epsilon}_i' \hat{\epsilon}_j}{n} = \frac{(y_i - \hat{y}_i)' (y_j - \hat{y}_j)}{n} \tag{3.28}
\]

where \(\hat{\sigma}_{ij}\) is the covariance of the \(i^{th}\) and the \(j^{th}\) equations if \(i \neq j\), and is the variance if \(i = j\). These values are the same for all \(n\) samples;\(^3\)

\(\hat{\epsilon}_i, \hat{\epsilon}_j\) are \(n\) by 1 matrices of the error term for the \(i^{th}\) and the \(j^{th}\) equations respectively;

\(n\) is the number of samples;

\(^2\)The notation used here is described on pages 14 and 15 of this thesis.

\(^3\)Zellner (1962) used the number of samples minus the number of exogenous variables in the system as the denominator. Revankar (1974) and other authors used \(n\). Other estimates of the degrees of freedom are given in Judge and others (1985), page 600.
\( y_i, y_j \) are \( n \times 1 \) matrices of the sample values for the endogenous variables of the \( i^{th} \) and the \( j^{th} \) equations, respectively;

\( \hat{y}_i, \hat{y}_j \) are \( n \times 1 \) matrices of the predicted values from the second OLS fit for the endogenous variables of the \( i^{th} \) and the \( j^{th} \) equations, respectively.

These variance estimates are then combined to obtain an estimate of the error covariance matrix for the system of equations. This matrix will be a block diagonal matrix with all elements of each diagonal in a block being equal (see equation 2.21).

The estimated error covariance matrix is then used to fit the system of equations for which the simultaneity bias was purged, using EGLS for the system of equations.

\[
y = \hat{Z}\Delta + E \tag{3.29}
\]

\[
\hat{\Delta}_{EGLS} = \left(\hat{Z}'\hat{\Omega}^{-1}\hat{Z}\right)^{-1}\hat{Z}'\hat{\Omega}^{-1}y \tag{3.30}
\]

\[
Var(\hat{\Delta}_{EGLS}) = \left(\hat{Z}'\hat{\Omega}^{-1}\hat{Z}\right)^{-1} \tag{3.31}
\]

where \( y \) is a \( gn \times 1 \) matrix of the endogenous variables;

\( \hat{Z} \) is a \( gn \times g((g - 1) + k) \) matrix of the RHS variables (endogenous variables substituted by predicted endogenous variables);

\( \Delta \) is a \( g((g - 1) + k) \times 1 \) matrix of the coefficients associated with the RHS variables. Coefficients are set to zero if the associated RHS variables do not affect the \( i^{th} \) endogenous variable; alternatively, the corresponding variable and its zero coefficient may be removed from the matrix;

\( E \) is a \( gn \times 1 \) matrix of the error terms associated with \( n \) samples of each of \( g \) endogenous variables;
\( \hat{\Delta}_{EGLS} \) is an estimate of the \( \Delta \) matrix;

\( \hat{\Omega} \) is a \( gn \) by \( gn \) matrix of the estimated error covariance matrix.

Because the simultaneity bias is removed in the first stage of 3SLS, estimated coefficients are consistent. These estimates are more efficient than 2SLS if contemporaneous correlation is significant and if the set of the RHS exogenous variables is different for each equation (Srivastava and Tiwari, 1978; Zellner and Theil, 1962), even for small sample sizes (Cragg, 1967). The 3SLS technique has also been shown to be less biased than 2SLS (Cragg, 1967) and therefore is less biased than OLS on individual equations. The estimated coefficients from 3SLS have been shown to be asymptotically normally distributed and asymptotically equivalent to the maximum likelihood estimators for systems of equations (full information maximum likelihood) if the error terms for each equation are considered to be normally distributed (Cragg, 1967).

For small samples, the iterated three stage least squares (I3SLS) may be used to obtain estimated coefficients and the error covariance matrix at the same time (Mikhail, 1975). Iterations to improve EGLS estimates for small samples in single equations and in SUR or simultaneous systems have been discussed by Dagenais (1978), Dhrymes (1971), Kmenta and Gilbert (1968), Magnus (1978), Madansky (1964), and Telser (1964). Generally, iterated EGLS is equal to the MLE estimates if error terms are normally distributed. The I3SLS technique is similar to the "zigzag" method proposed by Oberhoffer and Kmenta (1974) to obtain maximum likelihood estimates for single equations with non-iid error terms.

The 3SLS and I3SLS techniques were extended to systems with non-iid error terms for individual equations by adding more stages. Techniques from literature to estimate the error covariance matrix for SUR systems which have non-iid error terms for individual equations (Kmenta and Gilbert, 1970; Parks, 1967) were extended to simultaneous
systems in the same manner as SUR estimators were combined with 2SLS methods to obtain 3SLS methods. This extension of the SUR method assuming non-iid error terms to simultaneous systems is the basis for the MSLS technique derived as part of this research. The iterative analog has been termed iterated multistage least squares (IMSLS).

The MSLS procedure is therefore as follows:

1. A standard OLS technique is used to obtain predicted endogenous variables using all of the exogenous variables in the system (first stage equations). If the number of exogenous variables is high, some can be removed for easier computation with a loss in efficiency (Brundy and Jorgenson, 1971).

2. All endogenous variables on the RHS of equations in the system are replaced by their respective predicted values to purge simultaneity bias from the system. Simple or multiple linear regression is then used to obtain fitted second stage equations.\(^4\)

3. A consistent estimate of \( \Omega \) is obtained. The stages required to obtain a consistent estimate vary depending on the error structure of the system (Stages 2 and further).

4. The EGLS technique, extended to the system of equations, is used to obtain an estimate of \( \Delta \) as with Stage 3 of 3SLS.

5. To extend MSLS to IMSLS, a new value of \( \hat{\Omega} \) is computed using the estimated coefficients from the MSLS fit. The \( \hat{\Omega}_{\text{new}} \) matrix is restricted in the same way as the first \( \hat{\Omega} \) matrix. For instance, if the error covariance matrix for the first MSLS fit was considered to be block diagonal, as with 3SLS in which individual equations have iid error terms, the new estimated error covariance matrix would be also block

\(^4\)If error terms are non-iid, EGLS using an estimated error covariance matrix for each equation in the system can be used to obtain consistent estimates of the coefficients and their variances for the second stage equations. This is simply an extension of 2SLS to non-iid error terms. Further steps of the MSLS technique will only result in a gain in efficiency.
diagonal. The contemporaneous variances for the second iteration of a system of 
equations with a block diagonal error covariance matrix, are calculated as follows:

\[ \hat{\sigma}_{ij}^{3SLS} = \frac{\hat{\epsilon}_{i}^{3SLS} \hat{\epsilon}_{j}^{3SLS}}{n} = \frac{(Y_i - \hat{Y}_i^{3SLS})(Y_j - \hat{Y}_j^{3SLS})}{n} \]  
(3.32)

where \( \hat{\sigma}_{ij}^{3SLS} \) is the covariance of the \( i^{th} \) and the \( j^{th} \) equations if \( i \neq j \), and 
is the variance if \( i = j \). These values are the same for all \( n \) samples;
\( \hat{\epsilon}_{i}^{3SLS}, \hat{\epsilon}_{j}^{3SLS} \) are \( n \) by 1 matrices of the error terms for the \( i^{th} \) and the \( j^{th} \) 
equations, respectively, using the estimated coefficients from the 3SLS fit;
\( n \) is the number of samples;
\( y_i, y_j \) are \( n \) by 1 matrices of the sample values for the endogenous variables 
of the \( i^{th} \) and the \( j^{th} \) equations, respectively;
\( \hat{y}_i^{3SLS}, \hat{y}_j^{3SLS} \) are \( n \) by 1 matrices of the predicted values using the 
estimated coefficients from the first 3SLS fit for the endogenous variables 
of the \( i^{th} \) and the \( j^{th} \) equations, respectively.

6. A new EGLS fit for the system of equations is obtained by replacing the previous 
estimated error covariance matrix by the new \( \hat{\Omega} \).

\[ \hat{\Delta}_{new} = \left( \hat{Z}'\hat{\Omega}_{new}^{-1}\hat{Z} \right)^{-1} \hat{Z}'\hat{\Omega}_{new}^{-1} \hat{y} \]  
(3.33)

7. Steps 4 and 5 are then be repeated until convergence occurs. Because of the asymptotic properties of EGLS, this will only lead to improvements if the sample size is small.

If the error covariance matrix estimate (step 2) is consistent, the properties of generalized least squares (GLS) which uses the true error covariance matrix, can be assumed
for EGLS using large sample sizes (Judge et al., 1985, page 176). The estimated coefficients will be consistent, as will the estimate of the variances of the coefficients. Also, the distribution of the estimated coefficients will be asymptotically normal and confidence limits can be calculated using the normal probabilities (Maddala, 1974). For small sample sizes, the iteration should result in estimates which are quasi-maximum likelihood, or are MLE under the assumption that the error terms are normally distributed, as with the 3SLS procedure, assuming that the assumptions made concerning the error covariance matrix are correct. However, convergence of estimates by iterations is not assured.

To obtain a consistent estimate of the covariance matrix some assumptions about the nature of the covariance matrix must be made as it is not possible to estimate all of the elements of the error covariance matrix consistently (Judge et al., 1985, page 174). Also, many stages may be required and the stages required will differ depending on the assumptions made. For some non-iid error structures, procedures for obtaining consistent estimates of the error covariance matrix were not found in literature. For these error structures, existing techniques were modified and results published in literature were used as evidence that these modifications provide consistent estimates of the error covariance matrix.

3.2 Estimation of the Error Covariance Matrix

3.2.1 Autocorrelation and Contemporaneous Correlation

The error covariance matrix for a serially correlated single equation with homogenous variances (not independent, identically distributed) is the product of a scalar and a matrix given as follows (Judge et al., 1985, page 275):

$$ \Phi = \sigma_e^2 \Psi $$

(3.34)

where $\Phi$ is an $n$ by $n$ matrix of the error covariance;
\( \sigma^2 \) is the scalar multiplier;

\( \Psi \) is an \( n \) by \( n \) matrix.

If the error terms for an equation are iid, \( \Psi \) becomes the identity matrix \( I_n \). The assumption made in this thesis is that the relationship of the current error to the previous error remains the same for all of the error terms over the entire sample set (stationary process). For most forest inventory problems, the time from one observation to the next observation is usually quite long. Therefore, if serial correlation is present, the error for one observation is likely correlated with the observation immediately preceding it and is less correlated with the previous observations. This correlation was therefore assumed to follow a first order autoregressive process; the correlation between error terms declines geometrically with the distance between measurements. The assumption of first order correlation for most forestry inventory situations is therefore probably correct.

The \( \Phi \) matrix assuming a first order autoregressive process for a single equation is as follows:

\[
\Phi = \sigma^2_e \Psi = \frac{\sigma^2_e}{1 - \rho^2} \begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\
\rho & 1 & \rho & \ldots & \rho^{n-2} \\
\rho^2 & \rho & 1 & \ldots & \rho^{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1
\end{bmatrix} \tag{3.35}
\]

where \( \rho \) is the slope of \( \epsilon_m = \rho \epsilon_{m-1} + v_m \) and \( v_m \) are iid;

\( \sigma^2_e \) is equal to \( \sigma^2_v/(1 - \rho^2) \), where \( \sigma^2_v \) is the variance of \( v_m \).

The simplification made for this matrix is that there are time lapses of one unit between measurements. If the time lag differs from this, the elements of the matrix become \( \rho^s \) where \( s \) is the time between the error terms.\(^5\) The OLS estimation of an equation with this error covariance matrix using simple or multiple linear regression will result in an

\(^5\)See Judge and others (1985), pages 275 and 276 for more details.
unbiased estimate of the coefficients, but the usual estimate of the covariances of these coefficients will be biased. Also, the estimate of the coefficients will not be most efficient (Judge et al., 1985, page 278). The alternatives for single equations are to transform the equations and use simple or multiple linear regression or to estimate the error covariance matrix and use EGLS. These two methods have been proven equivalent by Jaech (1964).

Parks (1967) extended the first order autoregressive model to contemporaneously correlated systems of equations (SUR). In order to obtain an estimate of the error covariance matrix, \( \Omega \), for the system of equations, he performed the following steps:

1. For each equation, multiple linear regression analysis was performed. The coefficients are unbiased and, therefore, were used to obtain an estimate of the error for each sample for each equation \( \hat{\epsilon}_{im} \).

2. An estimate of \( \rho \) was obtained for each of the \( g \) equations of the system by regressing the current error term against the previous error term, resulting in \( n - 1 \) pairs of data, as follows:

\[
\hat{\rho}_i = \frac{\Sigma_{m=2}^{n} \hat{\epsilon}_{im} \hat{\epsilon}_{im-1}}{\Sigma_{m=2}^{n} \hat{\epsilon}_{im}^2 - 1}
\]  

(3.36)

where \( \hat{\epsilon}_{im} \) is the estimated residual for the \( i^{th} \) equation and the \( m^{th} \) observation;

\( \hat{\epsilon}_{im-1} \) is the estimated residual for the \( i^{th} \) equation and the \( m - 1^{th} \) observation.

3. The estimated values of \( \rho \) for each equation were used to obtain an estimate of \( P_i \).

\(^6\)For simultaneous systems, this first step would be performed using the second stage equations, and resulting coefficients would be biased, but consistent.
as follows:

\[ \hat{P}_i = \begin{bmatrix}
(1 - \hat{\rho}_i^2)^{-1/2} & 0 & 0 & \ldots & 0 \\
\hat{\rho}_i (1 - \hat{\rho}_i^2)^{-1/2} & 1 & 0 & \ldots & 0 \\
\hat{\rho}_i^2 (1 - \hat{\rho}_i^2)^{-1/2} & \hat{\rho}_i & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{\rho}_i^{n-1} (1 - \hat{\rho}_i^2)^{-1/2} & \hat{\rho}_i^{n-2} & \hat{\rho}_i^{n-3} & \ldots & 1
\end{bmatrix} \] (3.37)

where \( \hat{P}_i \hat{P}_i' = \hat{\Psi}_i / (1 - \hat{\rho}_i^2) \);

\( i \) refers to the \( i_{th} \) equation of the system;

\( \hat{\rho}_i \) is the slope of the regression line for \( \hat{\epsilon}_{im} \) with \( \hat{\epsilon}_{im-1} \) for the \( i^{th} \) equation.

The inverse of the \( \hat{P}_i \) matrix is as follows:

\[ \hat{P}_i^{-1} = \begin{bmatrix}
(1 - \hat{\rho}_i^2)^{1/2} & 0 & 0 & \ldots & 0 \\
-\hat{\rho}_i & 1 & 0 & \ldots & 0 \\
0 & -\hat{\rho}_i & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & -\hat{\rho}_i
\end{bmatrix} \] (3.38)

All elements of the diagonal are equal except the first element which estimates the relationship of \( \hat{\epsilon}_{i1} \) with \( \hat{\epsilon}_{i0} \) which cannot be exactly measured. Parks discussed the use of all diagonal elements equal to one. Maeshiro (1980) stated that the use of all diagonal elements equal to one results in a substantial reduction in efficiency, whereas Doran and Griffiths (1983) stated that little efficiency is lost. The loss in efficiency will depend on the how much the error terms of an equation are correlated, and also on the sample size.

4. Each equation was transformed using the estimated \( P_i^{-1} \) matrix, labelled as \( \hat{R}_i \).

\[ \hat{R}_i y_i = \hat{R}_i \hat{Z}_i \hat{\delta}_i + \hat{R}_i \hat{\epsilon}_i \] (3.39)
where $\hat{Z}_i$ is a matrix of predicted endogenous variables and exogenous variables (see equation 3.26). For the Parks study, all RHS variables were exogenous; 

$\delta_i$ is a matrix of coefficients for the $\hat{Z}_i$ with coefficients set to zero if the associated variables do not affect the $i^{th}$ equation.

The transformed regression equation has iid error terms, for large sample sizes.

5. Multiple regression was applied to each transformed equation and the estimated error terms were used to obtain the estimated contemporaneous variances from equation 3.28 for the following matrix.

\[
\hat{\Sigma} = \begin{bmatrix}
\hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \cdots & \hat{\sigma}_{1g} \\
\hat{\sigma}_{21} & \hat{\sigma}_{22} & \hat{\sigma}_{23} & \cdots & \hat{\sigma}_{2g} \\
\hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_{33} & \cdots & \hat{\sigma}_{3g} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{g1} & \hat{\sigma}_{g2} & \hat{\sigma}_{g3} & \cdots & \hat{\sigma}_{gg} 
\end{bmatrix}
\]  

(3.40)

6. The estimated error covariance matrix was obtained by the following product.

\[
\hat{\Omega} = \hat{P} \left( \hat{\Sigma} \otimes I_n \right) \hat{P}' = \begin{bmatrix}
\hat{\sigma}_{11}\hat{P}_1\hat{P}_1' & \hat{\sigma}_{12}\hat{P}_1\hat{P}_2' & \hat{\sigma}_{13}\hat{P}_1\hat{P}_3' & \cdots & \hat{\sigma}_{1g}\hat{P}_1\hat{P}_g' \\
\hat{\sigma}_{21}\hat{P}_2\hat{P}_1' & \hat{\sigma}_{22}\hat{P}_2\hat{P}_2' & \hat{\sigma}_{23}\hat{P}_2\hat{P}_3' & \cdots & \hat{\sigma}_{2g}\hat{P}_2\hat{P}_g' \\
\hat{\sigma}_{31}\hat{P}_3\hat{P}_1' & \hat{\sigma}_{32}\hat{P}_3\hat{P}_2' & \hat{\sigma}_{33}\hat{P}_3\hat{P}_3' & \cdots & \hat{\sigma}_{3g}\hat{P}_3\hat{P}_g' \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{g1}\hat{P}_g\hat{P}_1' & \hat{\sigma}_{g2}\hat{P}_g\hat{P}_2' & \hat{\sigma}_{g3}\hat{P}_g\hat{P}_3' & \cdots & \hat{\sigma}_{gg}\hat{P}_g\hat{P}_g' 
\end{bmatrix}
\]  

(3.41)

Parks proved that this estimator of $\Omega$ is consistent.

Kmenta and Gilbert (1970) tested the technique given by Parks (1967) and showed that other techniques are more efficient for small sample sizes. Two of the techniques tested were nonlinear techniques. One of the techniques, called ZEF-ZEF is a slight
modification of the Parks method which results in greater efficiency by obtaining an estimate of all of the $\rho_i$ using a simultaneous fit of all of the error prediction equations, rather than an OLS fit to obtain $\hat{\rho}_i$ for each error prediction equation separately.

The technique chosen to estimate the error covariance matrix for the MSLS technique was the Parks (1967) method with an added step from Kmenta and Gilbert (1970), between steps 2 and 3 as follows:

2a. The regression of $\hat{e}_{im}$ on $\hat{e}_{im-1}$ is used to obtain estimates of the following matrix.

$$\hat{\Sigma}_e = \begin{bmatrix}
\hat{\sigma}_{11e} & \hat{\sigma}_{12e} & \hat{\sigma}_{13e} & \cdots & \hat{\sigma}_{1ge} \\
\hat{\sigma}_{21e} & \hat{\sigma}_{22e} & \hat{\sigma}_{23e} & \cdots & \hat{\sigma}_{2ge} \\
\hat{\sigma}_{31e} & \hat{\sigma}_{32e} & \hat{\sigma}_{33e} & \cdots & \hat{\sigma}_{3ge} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{ge1} & \hat{\sigma}_{ge2} & \hat{\sigma}_{ge3} & \cdots & \hat{\sigma}_{gee}
\end{bmatrix}$$

(3.42)

$\hat{\sigma}_{ije}$ is calculated as follows:

$$\hat{\sigma}_{ije} = \frac{\hat{v}_i \hat{v}_j}{(n - 1)}$$

(3.43)

where $\hat{\sigma}_{ije}$ is the covariance of the $i^{th}$ and the $j^{th}$ error prediction equations if $i \neq j$, and is the variance if $i = j$. These values are the same for all $n$ samples;

$\hat{v}_i$ and $\hat{v}_j$ are $n$ by 1 matrices of the residual error from OLS fit of the error prediction equations for the $i^{th}$ and $j^{th}$ equations, respectively.

The error prediction equations are then fit simultaneously using EGLS for the system of error prediction equations where $\hat{\Omega}_e = (\hat{\Sigma}_e \otimes I_n)$. The estimated autocorrelation coefficients ($\hat{\rho}_i$) are then used to obtain the $\hat{P}_i$ matrices.

This technique results in a consistent estimate of $\Omega$ which is required so that the properties of GLS can be applied to EGLS asymptotically. The $P_i$ matrix can also be extended
to higher order correlations if these are found to be significant (see Ameniya, 1985, pages 164 to 170).

### 3.2.2 Heteroskedasticity and Contemporaneous Correlation

The error covariance matrix for heteroskedastic, contemporaneously correlated systems of equations without serial correlation (independent, not identically distributed) is block diagonal as shown in equation 2.22. The variance for an individual equation can be shown as $\Phi = \sigma^2\Psi$ where $\Psi$ is a diagonal matrix with unequal diagonal elements. To obtain a consistent estimate of the error covariance matrix, a consistent estimate of $\sigma^2$ and a consistent estimate of $\Psi$ are required.

A method for estimating the error covariance matrix for heteroskedastic, contemporaneously correlated systems of equations was not found in literature. Parks (1967) method for autocorrelated systems was therefore extended to heteroskedastic, contemporaneously correlated systems. The following steps were derived to obtain a consistent estimate of the error covariance matrix.

1. For each second stage equation, simple or multiple linear regression is used to obtain consistent estimates of the error terms for each sample as in step 1 for autocorrelated errors. The estimated coefficients will be consistent. If the system of equations is not simultaneous (stage 1 of MSLS procedure is not required), the estimated coefficients will be also unbiased.

2. The error terms, squared, are then graphed against the predicted values for the LHS endogenous variables using coefficients from step 1 for each equation as follows:

$$\hat{y}_{im} = \hat{Z}_{im}\hat{a}_i \quad (3.44)$$

$$u_{im} = \hat{\epsilon}_{im} = y_{im} - \hat{y}_{im} \quad (3.45)$$
where $u_{im}$ is $\epsilon_{im}$ for the $i^{th}$ equation and the $m^{th}$ observation;

$\hat{y}_{im}$ is the predicted endogenous variable from step 1 for the $i^{th}$ equation and the $m^{th}$ observation;

$\hat{Z}_{im}$ is the vector of predicted RHS endogenous variables and exogenous variables for the $i^{th}$ equation and the $m^{th}$ sample;

$\hat{\alpha}_i$ is the coefficient matrix from the step 1 simple or multiple linear regression fit of each equation.

3. Using the graph of estimated error squared versus predicted endogenous variables, an estimated functional form of the variance for each given sample (represented by the squared error) against the expected value of the endogenous variable given the RHS variables is chosen.

$$u^2_{im} = \hat{f}(\hat{y}_{im})$$

This equation may be fit using simple or multiple linear regression for each equation, resulting in a consistent estimate of the variances (Judge et al., 1985, p. 437) or all equations may be fit simultaneously using SUR as with step 2a of the process to obtain the error covariance matrix for autocorrelated systems.

4. The fitted equations are then used to obtain $\hat{\sigma}^2_{im} = \hat{u}^2_{im}$ to yield the following matrix.

$$W_i = \begin{bmatrix}
\hat{\sigma}_{i1e} & 0 & 0 & \ldots & 0 \\
0 & \hat{\sigma}_{i2e} & 0 & \ldots & 0 \\
0 & 0 & \hat{\sigma}_{i3e} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \hat{\sigma}_{ine}
\end{bmatrix}$$
Chapter 3. An Alternative Simultaneous Fitting Procedure

The inverse of this matrix is as follows:

\[
W_i^{-1} = \begin{bmatrix}
1/\hat{\sigma}_{i1c} & 0 & 0 & \ldots & 0 \\
0 & 1/\hat{\sigma}_{i2c} & 0 & \ldots & 0 \\
0 & 0 & 1/\hat{\sigma}_{i3c} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1/\hat{\sigma}_{ince}
\end{bmatrix}
\]  

(3.48)

5. The matrix \(W_i^{-1}\) is used to transform each of the \(g\) equations of the system resulting in the transformed equation as follows:

\[
W_i^{-1}y_i = W_i^{-1}\hat{Z}_i\delta_i + W_i^{-1}\epsilon_i
\]  

(3.49)

where \(\hat{Z}_i\) is a matrix of predicted endogenous variables and exogenous variables (see equation (3.26));

\(\delta_i\) is a matrix of coefficients for the \(\hat{Z}_i\) with coefficients set to zero if the associated variables do not affect the \(i^{th}\) equation.

The transformed regression equations have iid error terms, for large sample sizes.

6. Multiple linear regression is applied to each transformed equation and the estimated error terms from the weighted regressions are used to obtain the contemporaneous variances using equation 3.28.

\[
\hat{\Sigma} = \begin{bmatrix}
\hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \ldots & \hat{\sigma}_{1g} \\
\hat{\sigma}_{21} & \hat{\sigma}_{22} & \hat{\sigma}_{23} & \ldots & \hat{\sigma}_{2g} \\
\hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_{33} & \ldots & \hat{\sigma}_{3g} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{g1} & \hat{\sigma}_{g2} & \hat{\sigma}_{g3} & \ldots & \hat{\sigma}_{gg}
\end{bmatrix}
\]  

(3.50)
Chapter 3. An Alternative Simultaneous Fitting Procedure

7. The estimated error covariance matrix, \( \hat{\Omega} \) is obtained by the following product.

\[
\hat{\Omega} = W (\hat{\Sigma} \otimes I_n) W' = \\
\begin{bmatrix}
\hat{\sigma}_{11} W_1 W_1' & \hat{\sigma}_{12} W_1 W_2' & \hat{\sigma}_{13} W_1 W_3' & \cdots & \hat{\sigma}_{1g} W_1 W_g' \\
\hat{\sigma}_{21} W_2 W_1' & \hat{\sigma}_{22} W_2 W_2' & \hat{\sigma}_{23} W_2 W_3' & \cdots & \hat{\sigma}_{2g} W_2 W_g' \\
\hat{\sigma}_{31} W_3 W_1' & \hat{\sigma}_{32} W_3 W_2' & \hat{\sigma}_{33} W_3 W_3' & \cdots & \hat{\sigma}_{3g} W_3 W_g' \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\hat{\sigma}_{g1} W_g W_1' & \hat{\sigma}_{g2} W_g W_2' & \hat{\sigma}_{g3} W_g W_3' & \cdots & \hat{\sigma}_{gg} W_g W_g'
\end{bmatrix}
\] (3.51)

The estimate of \( \hat{\Omega} \) is the product of consistent estimates and is therefore also a consistent estimate. The assumption made in this explanation of how the estimated error covariance matrix is obtained was that the variances of the error terms for individual equations are related to the expected value of the endogenous variable given the exogenous variables. Other assumptions could be made in estimating the \( W_i \) matrix (see Judge et al., 1985, pages 424 and 425).

Duncan (1983) provided a proof that for heterogenous, contemporaneously correlated systems of equations, the SUR estimator is not most efficient. He suggested that an alternative estimate of the error covariance matrix be used to improve efficiency of the estimated coefficients. The modified Parks (1967) technique presented here should result in a improvement in efficiency for large sample sizes, if the assumptions made concerning the nature of the heteroskedasticity are correct.

### 3.2.3 Autocorrelation, Heteroskedasticity, and Contemporaneous Correlation

The error covariance matrix for single equations which are heteroskedastic and autocorrelated (neither independent nor identically distributed) is more complex in that the correlation of error terms is complicated by unequal variances. This can occur if data are
measured for different units over a period of time (cross sectional, time series data) such as with stem analysis data or with permanent sample plot data for forest inventories, or can occur for one unit measured over a long period of time during which climatic changes have influenced the relationship of error terms among measurements. These two situations present different problems and therefore different solutions were derived.

**Cross Sectional, Time Series Data**

Data which are collected by measuring each unit over a period of time are termed cross sectional, time series data. Data can be pooled and one equation can be fit to the pooled observations. In this case, the error term for a single equation can be divided into components and variance estimates can be obtained for (1) the variance for each unit across the time period (2) the variance for each time across the units and (3) the residual variance. The result is that the intercept of the equation varies over time and over units. The estimation of the error components has been attempted by many authors (e.g. Arora, 1973; Balestra and Nerlove, 1966; Wallace and Hussain, 1969) and is related to the use of indicator variables to indicate different time periods and units. However, time and unit differences may affect the slopes of the equation as well as the intercept.

An extension to slope coefficients has also been attempted by many authors (e.g. Hildreth and Houck, 1968; Swamy, 1970; Swamy and Mehta, 1977) and has been called random coefficient modelling. However, the estimation of random coefficients for data with unequal numbers of measurements within each unit is difficult. Also, even though estimates of variances can be found, the reasons for the time and cross sectional differences such as changes in climate, differences in genetics, and other factors are not explicitly explained by the equation.

For forestry problems, random coefficient modelling has been extended to random coefficient functions, in which parameters are predicted by other exogenous variables.
Chapter 3. An Alternative Simultaneous Fitting Procedure

resulting in a system of parameter prediction functions. First, an equation is fitted to each unit over the time periods. The estimated coefficients are then related to other exogenous variables using a set of parameter prediction equations. These parameter prediction equations are then fitted using simple or multiple linear regression on each equation. Ferguson and Leech (1978) argued that the traditional fitting of the parameter prediction equations did not account for the inter-relationships of the parameters, nor did it account for the error of the coefficients created by fitting one equation for each of the sample units. They proposed an alternative technique by estimating the error covariance matrix for the system of parameter prediction equations and by using this estimated matrix to find estimates of the coefficients of the system using EGLS. However, they did not attempt to account for the serial correlation in fitting an equation to each sample unit, resulting in a biased estimate of the variance of each coefficient. Also, they assumed that the errors for the parameter prediction equations were correlated, whereas the parameters themselves may be correlated resulting in a simultaneous set of parameter prediction equations.

For cross-sectional, time series data for forestry then, the procedure used by Ferguson and Leech (1978) could be modified by (1) accounting for serial correlation when fitting an equation to each sample unit through a transformation of variables or by estimating the error covariance matrix and using EGLS, and (2) identifying the parameter prediction equation as a system of simultaneous equations rather than a system of SUR equations. The MSLS procedure can be modified for a simultaneous, heteroskedastic system of parameter prediction equations as follows:

1. First, simple or multiple linear regression is used to fit a chosen equation to each of the sample units across time measurements. Equations would then be tested for serial correlation.
2. If serial correlation is found to be significant, an appropriate transformation followed by a new simple or multiple linear regression fit, or an estimation of the error covariance matrix and an EGLS fit of each equation separately is made. From this regression, consistent estimates of the coefficients and their associated variances are obtained.

3. The simultaneous system of parameter prediction equations is then chosen by using simple or multiple linear regression on each equation to determine which exogenous variables to include and to decide which parameters are strongly correlated.

4. To remove the simultaneity bias, each parameter is predicted from all of the exogenous variables in the system of parameter prediction equations. These predicted parameters are then substituted for parameters which appear on the RHS of equations in the system.

5. Multiple linear regression of each equation in the system is then performed.

6. The method proposed by Ferguson and Leech (1978) is then used to combine the variances from steps 2 and 5 into one matrix $\Omega$ and EGLS is used to fit the system of equations simultaneously.

**Single Unit Over A Long Period**

Autocorrelation and heteroskedasticity can also occur in individual units that are measured over a long period. In this case, the serial correlation between respective time measurements changes depending on the pair of measurements. This could occur in forestry as data from a permanent sample plot which has been measured many times over a long time period. In this case, climatic changes may have affected the relationships between time periods. In fact, the fitting of an equation to each unit as described
for cross sectional, time series data, may show error terms which are autocorrelated and heterogenous. Engle (1982) and Cragg (1982) proposed techniques to estimate coefficients under this type of error structure for individual equations. As an alternative, Gregoire (1987) suggested an initial transformation to remove either autocorrelation or heteroskedasticity.

For systems of equations which show this type of error structure, the following technique to estimate $\hat{\Omega}$ for step 2 of MSLS is proposed:

1. The transformation suggested by Cochrane and Orcutt (1949) is used to remove serial correlation, assuming first order autocorrelation. The transformation involves taking differences for all variables in the equation, meaning that one observation is lost.

2. Once serial correlation is removed, the steps given for heteroskedastic systems using the transformed data are followed.

The resulting estimated coefficients and their variances will be for the transformed data, but, for this transformation, they will also apply to the untransformed data. Also, because one observation is lost, the MSLS fit will not be asymptotically efficient. The loss in efficiency will depend on the number of samples, and the degree of the serial correlation.

### 3.3 Confidence Limits, Hypothesis Testing, and Prediction

The MSLS and 1MSLS procedures are extensions to 3SLS, which in itself is an extension to the EGLS procedure, applied to a system of equations. Because the EGLS procedure produces estimates of coefficients which are asymptotically normally distributed, for large
sample sizes, confidence limits for coefficients from MSLS can be calculated by the following:

\[ \hat{\delta}_{ir} \pm z_{\alpha/2} \hat{\sigma}_{\hat{\delta}_{ir}} \]  

(3.52)

where \( \hat{\delta}_{ir} \) is the \( r \)th estimated coefficient for the \( i \)th equation;

\( z_{\alpha/2} \) is the value from the standard normal probability distribution corresponding to the \( 1 - \alpha/2 \) percentile;

\( \hat{\sigma}_{\hat{\delta}_{ir}} \) is the estimated standard deviation for this coefficient which is the square root of the appropriate variance from the coefficient covariance matrix.

The coefficient covariance matrix is calculated as in equation 3.31.

For a single coefficient, the following hypothesis may be proposed.

\[ H_0 : \quad \hat{\delta}_{ir} = \delta_0 \]  

(3.53)

\[ H_1 : \quad \hat{\delta}_{ir} \neq \delta_0 \]  

(3.54)

where \( \delta_0 \) is the hypothesized value for the population value, \( \delta_{ir} \).

In order to test this hypothesis, a confidence interval for this coefficient could be calculated, and the hypothesis rejected if the confidence interval does not contain \( \delta_0 \). Alternatively, the following test statistic could be used.

\[ \text{test statistic} = \frac{\hat{\delta}_{ir} - \delta_0}{\hat{\sigma}_{\hat{\delta}_{ir}}} \]  

(3.55)

Under the assumption that the hypothesis is true, the test statistic is distributed asymptotically as the standard normal distribution. A one-sided hypothesis statement may also be tested.

If several coefficients of the system are to be tested as a group, the following hypothesis statement may be established.

\[ H_0 : \quad \mathbf{R}\Delta_{\text{reduced}} = \mathbf{r} \]  

(3.56)

\[ H_1 : \quad \mathbf{R}\Delta_{\text{reduced}} \neq \mathbf{r} \]  

(3.57)
where $\mathbf{R}$ is a $q$ by $K$ matrix;

$\mathbf{\Delta}_{\text{reduced}}$ is a matrix of coefficients in the system. Note that coefficients which have been constrained to be zero to represent RHS variables which do not affect the $i^{th}$ equation have been removed. The size of this matrix has therefore been reduced from $g(g - 1 + k)$ by 1 to $K$ by 1;

$r$ is a $q$ by 1 matrix;

$K$ is the number of coefficients in the system which were not constrained to be zero;

$q$ is the number of comparisons to be made.

Comparisons can be made between coefficients of different equations in the system. For instance, to test $H_0 : \delta_{11} = \delta_{21}$ or $H_0 : \delta_{11} - \delta_{21} = 0$, the following matrices would be derived.

$$R\hat{\mathbf{\Delta}}_{\text{reduced}} = r$$  \hspace{1cm} (3.58)

$$\begin{bmatrix}
\hat{\delta}_{11} \\
\hat{\delta}_{12} \\
\hat{\delta}_{21} \\
\hat{\delta}_{22} \\
\hat{\delta}_{23}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & -1 & 0 & 0
\end{bmatrix} = [0]$$  \hspace{1cm} (3.59)

The test statistic for 3SLS and therefore for MSLS is as follows:

$$\text{test statistic} = (R\hat{\mathbf{\Delta}}_{\text{reduced}} - r)' (R \text{Var} (\hat{\mathbf{\Delta}}_{\text{reduced}}) R)^{-1} (R\hat{\mathbf{\Delta}}_{\text{reduced}} - r)$$  \hspace{1cm} (3.60)

where $\text{Var} (\hat{\mathbf{\Delta}}_{\text{reduced}})$ is $\text{Var} (\hat{\mathbf{\Delta}})$ reduced by eliminating columns and rows which correspond to coefficients which were constrained to be zero.

This test statistic is asymptotically distributed as the $\chi^2$ distribution with $q$ degrees of freedom under the assumption that the hypothesis is correct (Judge et al., 1985, page 614).
Given a set of exogenous variables, the mean predicted LHS endogenous variables can be found using the coefficients from the MSLS fit.

1. The set of exogenous variables is put into the first stage equations from the first step of the MSLS procedure to obtain predicted endogenous variables and these are substituted for endogenous variables which appear on the RHS of equations in the system. The set of new exogenous variables and predicted RHS endogenous variables is used to create the following matrix.

\[
\hat{Z}_h = \begin{bmatrix}
\hat{Z}_{h1} & 0 \\
0 & \hat{Z}_{h2}
\end{bmatrix}
\]  

(3.61)

where \( \hat{Z}_{hi} \) is the matrix of RHS variables, new exogenous and predicted RHS endogenous, for the \( i^{th} \) equation.

2. The mean predicted LHS endogenous variables are then calculated by the following:

\[
\hat{y}_{|Z_h} = \hat{Z}_h \hat{\Delta}
\]  

(3.62)

where \( \hat{y}_{|Z_h} \) is the matrix of mean predicted LHS endogenous variables given the set of exogenous variables and predicted RHS endogenous variables.

The variance matrix of the mean predicted LHS endogenous variables can be calculated by the following:

\[
Var \left( \hat{y}_{|Z_h} \right) = \hat{Z}_h' Var \left( \hat{\Delta} \right) \hat{Z}_h
\]  

(3.63)

where \( Var \left( \hat{\Delta} \right) \) is calculated as in equation 3.31.

Confidence limits for a mean predicted LHS endogenous variable can be found by:

\[
\hat{y}_{|Z_h} \pm z_{\alpha/2} \hat{\sigma}_{\hat{y}_{|Z_h}}
\]  

(3.64)
As with the confidence limits for coefficients of the system, the normal distribution is used for \( \hat{y}_h \), because the asymptotic properties of EGLS are assumed to be appropriate for MSLS.
Chapter 4

Procedures for Comparison of Fitting Techniques

4.1 Selection of Equations for the Systems

In order to compare the alternative technique (MSLS or IMSLS) to the common approaches of (1) OLS applied to each equation and (2) the OLS fit of a composite model, first the equations of the system were chosen. Common linear equations were selected for each of the three forestry problems and, in situations where several equation models were available, each model was regressed using simple or multiple linear regression and a model was selected based on simplicity and a high coefficient of determination \(^1\) \((R^2\) value\)^2. Variables within the model were considered to be important to the model if the \(R^2\) change (the change in the \(R^2\) value due to the addition of that variable) was greater than 0.005. No attempt was made to relate this to a level of significance, because the assumption of normally distributed error terms was not made; therefore, the usual tests using the t or F distribution were not appropriate. The choice of equations was restricted to linear models to restrict the scope of this thesis; however, extensions of MSLS and IMSLS to nonlinear equations using generalized nonlinear regression should be possible.

---

\(^1\)Since the simple or multiple linear regression of equations of a simultaneous system of equations produces biased results, the coefficient of determination is only useful as a goodness-of-fit measure.

\(^2\)The notation, \(R^2\) value, will be used in this thesis to denote the coefficient of determination for both simple and multiple linear regression, although \(r^2\) is normally used for simple linear regression.
4.2 Obtaining the OLS, Composite, and MSLS Fits

Once the equations of the system were chosen, the residual errors for the simple or multiple linear regression fit of each equation were graphed and tested for (1) heteroskedasticity and (2) serial correlation. If either or both of these properties were noted for any equation, the OLS fit of this equation was appropriately modified so that the best OLS fit of each single equation of the system was obtained. The results of this fitting were then used as the single equation OLS fit for comparison with other techniques.

If possible, the equations of the system were combined into a composite model. OLS was then applied to this composite model and residuals were tested for heteroskedasticity and serial correlation. OLS was then modified to account for the properties of the error term and a new OLS for the composite model was fitted and used as a comparison with other techniques.

To obtain the fit for the MSLS method, the technique outlined in Chapter 3 was used as follows:

1. OLS was used to obtain a predicted value for each endogenous variable from all of the exogenous variables of the system (first stage equations). These predicted endogenous variables were then substituted for corresponding endogenous variables on the RHS of each of the equations of the system (Stage 1).

2. OLS was again used to fit each of the equations in the system once the predicted endogenous variables were substituted for the RHS endogenous variables (Stage 2, second stage equations).

3. The residuals from the second OLS fit were used to determine if heteroskedasticity or serial correlation were present in the system, because this second fit results in consistent estimates of the coefficients of the system.
4. Once the structure of the error covariance matrix was determined, an appropriate technique for estimating the error covariance matrix, $\Omega$, was chosen and used (Stages 3 and on). The number of steps required depended on the characteristics of the error structure. If neither heteroskedasticity nor serial correlation were present in the equations of the system, the error covariance matrix became block diagonal with all elements of the block being equal. MSLS is equal to 3SLS in this case, and 3SLS for large sample sizes or 13SLS for small sample sizes was used.

5. The estimated error covariance matrix, $\hat{\Omega}$, was then used to fit the system of equations simultaneously using EGLS for the system.

6. For small sample sizes, a new value for $\hat{\Omega}$ was then calculated using the estimated coefficients from the EGLS fit, and the process was iterated until convergence occurred. Convergence was considered to occur if each of the new coefficients was within 0.00005 of the corresponding previous coefficient as follows:

$$|\hat{\delta}_{il} - \hat{\delta}_{il(new)}| \leq 0.00005$$ (4.65)

where $\hat{\delta}_{il}$ is the $l^{th}$ coefficient for the $i^{th}$ equation.

If convergence did not occur, the first estimate (from step 5) was used. No attempt was made to try alternative nonlinear fitting algorithms.

The results from the MSLS (or IMSLS) were then used for the comparison with the more commonly used techniques. If a system of equations was found to be both heteroskedastic and autocorrelated, the MSLS steps were modified according to the description given in Chapter 3.

Once the set of coefficients, and their covariances were calculated for each of the techniques (OLS, composite model, and MSLS or IMSLS), the results of the three techniques were compared. Values which were compared among the three techniques included
goodness-of-fit measures, the traces of the estimated coefficient covariance matrices, and the estimated coefficients and their standard deviations. In addition, the techniques were ranked for (1) the amount of information about the system produced, (2) consistency of the estimators for the coefficients and for the variance of the coefficients as reported in literature, (3) ability to calculate confidence intervals, (4) asymptotic efficiency as reported in literature, (5) compatibility, and (6) ease of fit in terms of the number and difficulty of the steps required.

4.3 Tests for Heteroskedasticity and Serial Correlation

Although tests for serial correlation in systems of equations are available (Durbin, 1957; Harvey and Phillips, 1980), because the presence of heteroskedasticity or serial correlation in any equation will affect the error covariance matrix of the system of equations, tests for single equations were chosen.

For serial correlation, the error terms were ordered in time sequence by age if available (i.e. tree or plot age), or by the magnitude of the predicted endogenous variable on the LHS, from the smallest to largest value. A graph of the $\hat{\epsilon}_{im}$ with the $\hat{\epsilon}_{im-1}$ and simple linear regression analysis were done. If the coefficient of determination from this regression was greater than 0.005, the Durbin and Watson bounds test (1951) was used; Epps and Epps (1977) found this to be reliable even if heteroskedasticity is present. However, the bounds test can only be used if an intercept item is included in the equation. Also, Harvey and Phillips (1980) found that the zone representing inconclusive results is large if the number of predictor variables is large. For forestry problems, the number of predictor variables is generally small. The test statistic for the bounds test is as follows:

$$d_i = \frac{\sum_{m=2}^{n} (\hat{\epsilon}_{im} - \hat{\epsilon}_{im-1})^2}{\sum_{m=1}^{n} \hat{\epsilon}_{im}^2}$$

(4.66)

where $\hat{\epsilon}_{im}$, $\hat{\epsilon}_{im-1}$ are the estimated errors from the simple or multiple linear regression fit
of the \(i^{th}\) equation of the system, once simultaneity bias is removed.

The tables for the Durbin and Watson bounds test (1951) were extended to more RHS variables and to larger and smaller sample sizes by Savin and White (1977) and these more extensive tables were used. However, because the bounds test has a region of inconclusiveness, and the tables are only available for up to a sample size of 200, a second test used by Newberry and Burkhart (1986), called the runs test, was performed if the results of the bounds test were inconclusive or if the sample size was large. For the runs test, each ordered \(\hat{\varepsilon}_im\) and \(\hat{\varepsilon}_{im-1}\) pair was given a "+" if the difference value is positive. If the difference was negative, a "−" is assigned to the pair. The sign test (Conover, 1980) was then used to determine if the number of positive pairs is significantly large (positive serial correlation), or alternatively, if the number of positive pairs is significantly small (negative serial correlation) using a one-sided test. A two-sided test can be used to test for the presence of serial correlation, either positive or negative simultaneously.

To examine the error terms for heteroskedasticity, a graph of \(\hat{\varepsilon}_im^2\) versus \(\hat{y}_{im}\) and simple linear regression were done. If the coefficient of determination from this regression was greater than 0.005, further testing was done; otherwise, the error terms were considered homoskedastic. If a test was deemed necessary and if no serial correlation was found, the Goldfeld and Quandt test for heteroskedasticity (1965) was used. The data were first ordered by the predicted endogenous variable, from smallest to largest if the variances were considered nondecreasing, and from largest to smallest, if the variances were considered to be nonincreasing. The ordered data were then divided into three groups, and a linear regression was performed for the group with the lower variances and for the group with the higher variances. The ratio of the mean squared error for the regression of data with the higher variances, to the mean squared error for the regression of data with the lower variances was calculated as the test statistic. This test assumes that the variance is monotonically increasing or decreasing, which is likely for forestry situations.
If serial correlation was found, the data were first transformed to remove serial correlation (Cochrane and Orcutt, 1949), because the Goldfeld and Quandt test was found to be in error if serial correlation is present (Epps and Epps, 1977). The Goldfeld and Quandt test was then applied to the transformed data.

For all tests, an alpha level of 0.05 was used.

4.4 Criteria for Comparison of the Three Techniques

4.4.1 Goodness-of-fit Measures

For comparison of the fit of the three techniques for the sample data, the goodness-of-fit measures were calculated. Because the OLS technique minimizes the sum of the squared error, goodness-of-fit measures based on the square or absolute value of the error terms will be best for the OLS fit. However, these measures are commonly used in forestry to indicate how well the fitted equation relates to the sample data; therefore, this comparison was made. The goodness-of-fit measures do not indicate whether the estimated coefficients are unbiased, however.

1. For the OLS and MSLS fitting techniques, the Fit Index (F.I.) was calculated for each of the $g$ equations of the system.

$$F.I._i^2 = 1 - \frac{\sum_{m=1}^{n} (y_{im} - \bar{y}_{im})^2}{\sum_{m=1}^{n} (y_{im} - \bar{y}_i)^2}$$  \hspace{1cm} (4.67)

where $n$ is the number of samples;

$y_{im}, \bar{y}_{im}$ are values for the $m^{th}$ sample of the $i^{th}$ equation for the endogenous and predicted endogenous variables, respectively;

$\bar{y}_i$ is the average of the sample values for the $i^{th}$ endogenous variable.
The F.I. is expected to be highest for the OLS fit of each of the equations, because
the OLS fit minimizes the unweighted error terms; whereas, the MSLS fit minimizes
the weighted error terms.

To compare the composite model to the alternative methods for goodness-of-fit, the
F.I. was calculated for the composite model and compared to the F.I. calculated
for each of the other two techniques for the corresponding endogenous variable.

2. The Mean Absolute Difference was calculated for each equation over the range of
the sample data to compare the size of the differences between the OLS and the
MSLS fits. The data were evenly divided into five classes based on the value for
the endogenous variable and the mean difference was calculated for each class.

\[
M.A.D._{i,l} = \frac{1}{c} \sum_{f=1}^{c} |y_{lf} - \hat{y}_{lf}|
\]

(4.68)

where \( M.A.D._{i,l} \) is the mean absolute deviation for the \( i^{th} \) endogenous variable in
the \( l^{th} \) class;

\( c \) is the number of observations in the \( l^{th} \) class, approximately equal to \( n/5 \)
for all equations and all classes.

To compare the composite model to the other fitting techniques, the M.A.D. was
calculated for each class of the endogenous variable which appears as a LHS variable
in the composite model.

3. The Mean Difference (M.D.) was calculated for each equation over the range of
the sample data to compare the direction of differences between the OLS and the
MSLS fits. The data were evenly divided into five classes based on the value for
the endogenous variable and the mean difference was calculated for each class.

\[
M.D._{i,l} = \frac{1}{c} \sum_{f=1}^{c} (y_{lf} - \hat{y}_{lf})
\]

(4.69)
where $M.D._{ii}$ is the average deviation for the $i^{th}$ endogenous variable in the $l^{th}$ class;

$c$ is the number of observations in the $l^{th}$ class, approximately equal to $n/5$ for all equations and all classes.

To compare the composite model to the alternative techniques, the M.D. was calculated for each of the three techniques for the variable which appears as the LHS variable in the composite model.

4.4.2 Trace of the Estimated Coefficient Covariance Matrix

The OLS fit of simultaneous equations results in biased and inconsistent estimates of the coefficients. The comparison of the traces of the estimated coefficient covariance matrix (sum of the estimated variances of the coefficients) from the OLS fit to that from the MSLS fit will not, therefore, indicate the relative efficiency. However, because the use of the estimated coefficient covariance matrix from the OLS fit is common, and the differences in the estimated matrices are of interest, a comparison was made. Also, this comparison was used to examine the magnitude of the differences and, therefore, to indicate that the use of the OLS fit for testing hypothesis statements and for calculating confidence limits for simultaneous systems, can be quite incorrect. Initially, the traces for the complete system of equations were compared, and then the traces of each submatrix, corresponding to a single equation, were compared for the two techniques.

For comparison of the composite model to the OLS, the trace of the estimated covariance matrix for the fit of the composite model was compared to the trace of the estimated covariance matrix for the OLS fit of the equation having the same endogeneous variable on the LHS as the composite model. Similar trace values were calculated to compare the MSLS fit to the fit of the composite model.
4.4.3 Table of Estimated Coefficients and Standard Deviations

Because the trace of an estimated coefficient covariance matrix may mask some of the
differences for single coefficients, a table of the estimated coefficients and their standard
deviations was compiled. Because the composite model fit does not necessarily have the
same variables, only the OLS, MSLS, IMSLS fits were compared.

4.4.4 Ranking for Other Features

An overall ranking system has been used by several authors including Amateis and
others (1984). The ranking system used here is simple in that a higher rank indicates
that the technique has more of the desirable criteria of fitting techniques. The ranks were
assigned as follows:

1. The amount of information given.
   Rank = 1, if the technique estimates one variable only;
   Rank = 2, if all endogenous variables are estimated.

2. The consistency of the coefficient and covariance matrix estimates as reported in
   literature.
   Rank = 1, if neither is consistent;
   Rank = 2, if both are consistent.

3. Ability to calculate confidence limits (requires a consistent estimate of the covari-
   ance matrix and a known probability distribution).
   Rank = 1, not able;
   Rank = 2, can be calculated.
4. Asymptotic efficiency as reported in literature.
   Rank=1, larger variance;
   Rank=2, smaller variance.

5. Compatibility across equations.
   Rank=1, not compatible;
   Rank=2, compatible.

6. Ease of fit.
   Rank=1, most difficult;
   Rank=2, medium difficulty;
   Rank=3, least difficult.

The ranks were assigned and summed for each of the three techniques used. For each criterion, equal weight was assigned, except that for the last item three ranks were assigned. The ranks could have been given other weights depending on the item, but the weights given are largely a function of the researcher or his organization, and so the weighting of ranks would likely change depending on the user of the information. For this thesis, the equal weighting was chosen; information was provided so that other weights could be assigned and a new overall rank calculated for each fitting technique if desired.
Chapter 5

Application 1: Tree Volume Estimation

5.1 Introduction

The estimation of gross tree volume has been well researched and many methods and models have been employed. Models have been developed for the estimation of total volume, which includes volume of the main bole of the tree from ground to tree tip, and for merchantable volume which includes volume for the merchantable part of the main bole only. Two main types of models have been developed. The first type uses a system of equations to estimate total volume, and to estimate the ratio of merchantable volume to total volume (volume ratio), and then merchantable volume is calculated from the volume ratio and the total volume. The second type of model for estimating tree volume uses a mathematical equation to represent the shape of the tree bole (taper), and then integration is used to estimate volume for the whole bole or for any merchantable part. To limit the scope of this thesis, the examination of tree volume estimation was restricted to the first type of model, although the second type of volume estimation models using a taper function can also be considered a system of equations (Reed and Green, 1984).

5.2 Preparation of Data

Sectioned tree data for 1818 pine trees in Alberta were obtained from the Alberta Forest Service (AFS). Trees were sampled by AFS personnel by selecting particular individuals with desired traits, or by selecting plots within stands and sectioning all of the trees in
the selected plot. Trees were sectioned at 0.3 metres above ground (stump height), 1.3 metres above ground (breast height), 2.8 metres above ground (2.5 metres above stump height), and subsequently at 2.5 metre intervals. Sections were further divided if decay was found within the section. For each section, the diameter inside and outside bark at the top of the section were measured and the section lengths were recorded. The number of tree rings and the dimensions of any decay were also measured at the top of each section. More details about the tree sectioning technique can be found in *Alberta Phase 3 Forest Inventory: Tree sectioning manual* (Anon., 1985).

Data were collected throughout the province. To limit these data to lodgepole pine (*Pinus contorta* var. *latifolia* Engelm.), data from Northeastern Alberta were removed, because most of the trees in this region are considered to be jack pine (*Pinus banksiana* Lamb.). Also, trees which were forked or had broken tops were deleted. Table 5.2 is a distribution of the remaining 1097 trees.

For each tree, total volume from ground to tip was calculated by assuming that the first section has a cylindrical shape (from ground to 0.3 m above ground), the top section has a conical shape, and the remaining sections have a paraboloid frustum shape. Merchantable volume was also calculated for each tree from a 0.3 metre stump height above ground, to a 7.0, 10.0, 13.0, and 15.0 centimetre top diameter inside bark (top dib). In addition, tree height (sum of the section lengths), merchantable length (the length of the merchantable part of the stem), dbh, and stump dib (diameter inside bark at the top of the first section) were calculated. Each tree therefore represented four merchantability standards with total volume included as another merchantability standard (volume from a 0.00 metre stump height to 0.0 cm top dib). Because these five merchantable volumes per tree do not represent independent data, only one of the five merchantable volumes was retained for each tree. To determine which of the five merchantable volumes to retain on a tree, a systematic selection was performed by selecting the first merchantable
Table 5.2: Distribution of Selected Trees by Height and Dbh Classes

<table>
<thead>
<tr>
<th>height in metres</th>
<th>1.30 to 5.00</th>
<th>5.01 to 10.00</th>
<th>10.01 to 15.00</th>
<th>15.01 to 20.00</th>
<th>20.01 to 25.00</th>
<th>25.01 to 30.00</th>
<th>30.01 to 35.00</th>
<th>35.01 to 40.00</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbh in cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1 to 5.0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>5.1 to 10.0</td>
<td></td>
<td>21</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>10.1 to 15.0</td>
<td></td>
<td>15</td>
<td>104</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>139</td>
</tr>
<tr>
<td>15.1 to 20.0</td>
<td></td>
<td>7</td>
<td>89</td>
<td>152</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td>270</td>
</tr>
<tr>
<td>20.1 to 25.0</td>
<td></td>
<td>1</td>
<td>25</td>
<td>138</td>
<td>73</td>
<td>3</td>
<td></td>
<td></td>
<td>240</td>
</tr>
<tr>
<td>25.1 to 30.0</td>
<td></td>
<td>3</td>
<td>48</td>
<td>99</td>
<td>29</td>
<td>1</td>
<td></td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>30.1 to 35.0</td>
<td></td>
<td>9</td>
<td>59</td>
<td>56</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>127</td>
</tr>
<tr>
<td>35.1 to 40.0</td>
<td></td>
<td>3</td>
<td>21</td>
<td>40</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td>40.1 to 45.0</td>
<td></td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>45.1 to 50.0</td>
<td></td>
<td>1</td>
<td></td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>50.1 to 55.0</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>55.1 to 65.0</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4</strong></td>
<td><strong>44</strong></td>
<td><strong>233</strong></td>
<td><strong>371</strong></td>
<td><strong>279</strong></td>
<td><strong>148</strong></td>
<td><strong>17</strong></td>
<td><strong>1</strong></td>
<td><strong>1097</strong></td>
</tr>
</tbody>
</table>
volume (0.00 stump height to 0.0 cm top db) for the first tree in the sample, the second merchantable volume (0.30 stump height and 7.0 cm top db) for the second tree in the sample, and so on. The result was that approximately one fifth of the sample trees represented each of the five merchantability classes.

Because some of the sectioned trees were sampled from plots, the tree data were not considered independent. Also, if all 1097 sample trees were used in the analysis of systems of equations, the size of the error covariance matrix would be $g \times 1097$ by $g \times 1097$. The inverse of this error covariance matrix must be calculated for the MSLS procedure, and the calculation would be difficult for an array of this size. To reduce the dependence of trees sampled within plots, and also to reduce the size of the error covariance matrix, a random sample was selected from the 1097 trees. Initially, 500 trees were selected, but, because only seven megabytes of computer memory for running computer programs were available, difficulties with inverting the $g \times 500$ by $g \times 500$ error covariance matrix were encountered. Therefore, a second sample of 100 trees was selected for the analysis. The sample data were graphed and no outliers were found. The distribution of these 100 sample trees is presented in Table 5.3.¹

5.3 Model Selection

The equation to estimate total volume was restricted to standard volume functions which predict total volume as a function of total height and dbh. The models selected for possible inclusion into the system of equations were as follows:

1. A nonlinear model proposed by Schumacher and Hall (1933) was selected for testing.

$$total \ volume = \beta_0 \ dbh^{\beta_1} \ height^{\beta_2} \ error_1$$

(5.70)

¹A copy of the data selected for each application presented in this thesis can be obtained on diskette or tape by contacting the author.
### Table 5.3: Distribution of 100 Sample Trees

<table>
<thead>
<tr>
<th>dbh in cm</th>
<th>height in metres</th>
<th>1.30 to 5.00</th>
<th>5.01 to 10.00</th>
<th>10.01 to 15.00</th>
<th>15.01 to 20.00</th>
<th>20.01 to 25.00</th>
<th>25.01 to 30.00</th>
<th>30.01 to 35.00</th>
<th>35.01 to 40.00</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 to 5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5.1 to 10.0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>10.1 to 15.0</td>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>15.1 to 20.0</td>
<td>7</td>
<td>14</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>20.1 to 25.0</td>
<td>2</td>
<td>14</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>25.1 to 30.0</td>
<td></td>
<td>4</td>
<td>10</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>30.1 to 35.0</td>
<td></td>
<td></td>
<td>7</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>35.1 to 40.0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>40.1 to 45.0</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>45.1 to 50.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>50.1 to 55.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>55.1 to 65.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0</td>
<td>3</td>
<td>21</td>
<td>33</td>
<td>30</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>
where total volume is the volume of the main bole of the tree from ground to tree tip;

dbh is the diameter inside bark measured at 1.3 metres, called diameter at breast height;

height is the total tree height from ground to tree tip;

$\beta_0, \beta_1, \beta_2$ are coefficients to be estimated;

$\text{error}_1$ is the error term.

Using logarithms, the model can be transformed to a linear model, and then estimated coefficients can be obtained using multiple linear regression. This model was found to be the most applicable to data from Alberta (LeMay, 1982). However, results are in terms of logarithms and must be converted to the original units.

2. A simple linear model proposed by Spurr (1952) was shown by LeMay (1982) to be also quite accurate for Alberta data.

$$\text{total volume} = \beta_3 + \beta_4 \text{dbh}^2 \text{height} + \text{error}_2 \quad (5.71)$$

where $\beta_3, \beta_4$ are coefficients to be estimated;

$\text{error}_2$ is the error term.

Spurr’s model is based on the mathematical relationship of area at the base and height to volume.

3. Honer (1965) proposed a transformation of total volume resulting in the following model.

$$\frac{\text{dbh}^2}{\text{total volume}} = \beta_5 + \frac{\beta_6}{\text{height}} + \text{error}_3 \quad (5.72)$$
where $\beta_5, \beta_6$ are coefficients to be estimated;

$error_3$ is the error term.

4. Another simple linear model was identified as follows:

$$Total\ Volume = \beta_7 + \beta_8 dbh + \beta_9 dbh^2 + \beta_{10} height + error_4$$ (5.73)

where $\beta_7, \beta_8, \beta_9, \beta_{10}$ are coefficients to be estimated;

$error_4$ is the error term.

This model is simple in that total volume, total height and dbh are all untransformed.

Generally, the variance of the error terms for total volume linear models are heteroskedastic. The logarithmic transformation required to linearize the first model removes this heteroskedasticity.

For each of these models, simple or multiple linear regression was used to estimate coefficients. Spurr's model was selected for the system of equations to calculate volume, because the model is simple, logically based on the calculation of volume using geometric formulae, and the fit statistics were high (coefficient of determination ($R^2$) value of 0.9881). \(^2\) The Schumacher and Hall, and the Honer models require a transformation of the total volume variable. Since transformations of the endogenous variables appearing on the LHS of equations in the systems might cause unnecessary complications, these models were rejected. The remaining linear model (equation 5.73) was not selected as the $R^2$ value for this model (0.9827) was less than that of the Spurr model.

For the prediction of merchantable volume, two equations are used. An equation to estimate the volume ratio is paired with an equation which calculates merchantable

\(^2\)Comparison of the coefficient of determination as a goodness-of-fit measure was made; however, for the Schumacher and Hall, and the Honer models, this coefficient of determination is based on the transformation of volume.
volume from the estimated total volume and volume ratio. Honer (1964, 1967) proposed several equations to estimate volume ratio.

\[
VR = \beta_{11} + \beta_{12} \left( \frac{hm}{height} \right) + \beta_{13} \left( \frac{hm}{height} \right)^2 + error_5 \tag{5.74}
\]

\[
VR = \beta_{14} + \beta_{15} \left( \frac{top\ dib}{dbh} \right)^2 + \beta_{16} \left( \frac{top\ dib}{dbh} \right)^4 + error_6 \tag{5.75}
\]

\[
VR = \beta_{17} + \beta_{18} \left( 1 + \frac{hs}{height} \right) \left( \frac{top\ dib}{dbh} \right)^2 + \beta_{19} \left( 1 + \frac{hs}{height} \right) \left( \frac{top\ dib}{dbh} \right)^2 + error_7 \tag{5.76}
\]

\[
VR = \beta_{20} + \beta_{21} \frac{top\ dib}{dbh} + \beta_{22} \left( 1 - \frac{hm}{height} \right)^2 + error_8 \tag{5.77}
\]

where \( VR \) is the volume ratio;

\( hm \) is merchantable height, or the merchantable length plus the stump height;

\( top\ dib \) is the top diameter inside bark for the merchantability standard;

\( hs \) is the stump height, 0.0 for the first merchantability standard (total volume), and 0.30 for the remaining four merchantability standards;

\( \beta_{11} \) through \( \beta_{22} \) are coefficients to be estimated;

\( error_5 \) through \( error_8 \) are the error terms.

Honer (1964) used dbh measured inside bark, whereas Honer (1967) used dbh measured outside bark. For this thesis, dbh measured outside bark was used as this measure can be easily obtained on standing trees.

Each model was fitted using multiple linear regression. The highest \( R^2 \) value (0.9966) was obtained for Honer's fourth model, but this model is based on merchantable height and does not reflect changes in stump height. For the first of Honer's models, the \( R^2 \) value was 0.9962, and for the second model, the \( R^2 \) value was 0.8934. As with the fourth model, the first and second models do not reflect changes in stump height. The remaining model, equation 5.76, does reflect changes in both stump height and in the top dib, but
the $R^2$ value for this model was only 0.8943, somewhat lower than that for the fourth model.

To reflect the changes in stump height, and to retain the high fit statistics of the fourth model, the following model was derived.

$$VR = \beta_{23} + \beta_{24} \left( 1 - \frac{ml}{height} \right)^2 + error_9$$ (5.78)

where $ml$ is the merchantable length, which is the merchantable height minus the stump height;

$\beta_{23}, \beta_{24}$ are coefficients to be estimated;

$error_9$ is the error term.

The top $dib/dbh$ ratio was removed from the model as the inclusion of this ratio to the model resulted in a change of the $R^2$ value of only 0.0006. The simple linear regression fit of this simple model produced a high $R^2$ value (0.9960) and was chosen for the system of equations to estimate volume.

The total volume equation model relies on the measurement of tree dbh and height. Often height is not measured and rather is predicted from dbh. In British Columbia, the B.C. Ministry of Forests and Lands has approved the following height models (Watts, 1983) for use.

$$height = \beta_{25} + \beta_{26} dbh + \beta_{27} dbh^2 + error_{10}$$ (5.79)

$$height = 1.3 + \beta_{28} dbh + \beta_{29} dbh^2 + error_{11}$$ (5.80)

$$height = \beta_{30} + \frac{\beta_{31}}{dbh} + \beta_{32} dbh + error_{12}$$ (5.81)

$$height = 1.3 + \frac{\beta_{33} dbh}{dbh + 1} + \beta_{34} dbh + error_{13}$$ (5.82)

where $\beta_{25}$ through $\beta_{34}$ are coefficients to be estimated;

$error_{10}$ through $error_{13}$ are error terms.
Chapter 5. Application 1: Tree Volume Estimation

The equation used in Alberta is as follows (Edwards, 1987):

\[
\text{height} = \beta_{35} e^{\frac{dbh}{\beta_{36}}} \text{error}_{14}
\]  

(5.83)

where \( \beta_{35}, \beta_{36} \) are coefficients to be estimated;

\( \text{error}_{14} \) is the error term.

This model can be transformed to a linear model using logarithms.

Multiple linear regression was used to fit each of these equations. The first model, the paraboloid model, was selected as the height model to be included in the system of equations based on simplicity and an \( R^2 \) value of 0.7699. The second and fourth models were rejected as the intercepts are conditioned to 1.3 metres, and the effect of this conditioning, if the true intercept is not 1.3, would be that estimated coefficients are biased. The fifth model, the Alberta model, was rejected as a transformation of height is required to obtain a linear model and this complication was deemed unnecessary as results from other nontransformed models were superior. The remaining model, the hyperbola (equation 5.81), was rejected in that the \( R^2 \) value was 0.7636, slightly lower than the chosen model.

The selected volume ratio model requires the ratio of merchantable length to total height (height ratio). Again, since the height and merchantable length are likely not measured, a model for the estimation of the height ratio was required. Several linear models were fitted based on transformations of top dib, dbh, and stump height. The following model was selected.

\[
HR = \beta_{37} + \beta_{38} \left( \frac{\text{top dib}}{dbh} \right)^2 + \beta_{39} h s + \text{error}_{15}
\]  

(5.84)

where \( HR \) is the height ratio;

\( \beta_{37}, \beta_{38}, \beta_{39} \) are coefficients to be estimated;

\( \text{error}_{15} \) is the error term.
Because \textit{top dib} ranged from 0.00 through 15.0 centimetres, for small trees, the specified \textit{top dib} was sometimes larger than \textit{dbh}. To restrict the \textit{top dib/dbh} ratio, the ratio was allowed a maximum value of \textit{stump dib/dbh} where \textit{stump dib} is the diameter inside bark measured at stump height. The $R^2$ value for this model was 0.9722. Using a stepwise regression procedure, other transformations of the \textit{top dib/dbh} ratio, such as the first and third power, entered into the equation before $hs$; however, when these other transformations of the diameter ratio were included in the equation, the inclusion of $hs$ in the model resulted in an $R^2$ value change of less than 0.005. Since the effect of changes in the stump height on the height ratio was considered important, the equation presented above was selected for the system of equations.

The chosen system of equations for estimating tree volume was the following:

\begin{align*}
\text{height} &= \delta_{11} + \delta_{12}\text{dbh} + \delta_{13}\text{dbh}^2 + \epsilon_1 \tag{5.85} \\
\frac{ml}{\text{height}} &= HR = \delta_{21} + \delta_{22} \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 + \delta_{23}hs + \epsilon_2 \tag{5.86} \\
\text{total volume} &= \delta_{31} + \delta_{32}\text{dbh}^2\text{height} + \epsilon_3 \tag{5.87} \\
\frac{\text{merch. volume}}{\text{total volume}} &= VR = \delta_{41} + \delta_{42}(1 - HR)^2 + \epsilon_4 \tag{5.88} \\
\text{merch. volume} &= \text{total volume} \times VR \tag{5.89} \\
ml &= \text{height} \times HR \tag{5.90} \\
\text{merch. height} = hm &= ml + hs \tag{5.91}
\end{align*}

The system is therefore composed of seven equations, and estimates of coefficients are needed for four of these equations.
5.4 Ordinary Least Squares Fit

5.4.1 Unweighted Simple or Multiple Linear Regression

The estimated coefficients for the chosen equations using unweighted multiple linear regression to fit each equation were as follows:

\[
\text{pred. height} = 0.584579 + 1.071239 \text{dbh} - 0.009644 \text{dbh}^2 \\
\text{pred. HR} = 1.000000 - 0.912563 \left( \frac{\text{top dbh}}{\text{dbh}} \right)^2 - 0.226424 \text{hs} \\
\text{pred. total volume} = 0.011944 + 3.55499 \times 10^{-5} \text{dbh}^2 \text{height} \\
\text{pred. VR} = 0.990691 - 0.986292 (1 - \text{HR})^2
\]  

(5.92)  
(5.93)  
(5.94)  
(5.95)

The estimated error terms using these unweighted fitted equations were used to test for serial correlation and heteroskedasticity.

5.4.2 Testing for iid Error Terms

To check for serial correlation in each equation, the estimated error terms were sorted by the number of annual rings counted at stump height on each tree (age), and a graph\(^3\) of the current error term versus the previous error term was obtained. Simple linear regression was performed for the current error term with the previous error term, and all four \(R^2\) values (one for each of the four equations) were less than 0.005. Also, for an alpha level of 0.10 (alpha of 0.05 for positive correlation and an alpha of 0.05 for negative correlation), the Durbin and Watson (1951) statistic was not significant for any of the equations and so the error terms were considered to be independent for each of the four equations of the system.

For the height model, the graph of the estimated error versus the predicted height

\(^3\)Graphs used in this research are not presented in this thesis, because of their large number; graphs are on file and copies can be made available.
did not indicate heteroskedasticity, but a simple linear regression of the estimated error squared with predicted height resulted in an $R^2$ value of 0.01158. The Goldfeld and Quandt (1965) test for heteroskedasticity was therefore used to test the height model. The estimated error terms were ordered by the predicted values of the height variable from smallest to largest, because the variances were considered to be nondecreasing. Using the ordered data, multiple linear regression was used to fit the height model using the first 40 samples only (small estimated error terms) and then using the last 40 samples (large estimated error terms) only. The 20 samples representing the center of the ordered estimated error terms were not included in either regression. The test statistic was calculated as follows:

$$\text{test statistic} = \frac{\text{mean squared error}_2}{\text{mean squared error}_1} = \frac{7.32302}{5.60853} = 1.3057 \quad (5.96)$$

where mean squared error 2, mean squared error 1 are the mean squared errors for the regression of the larger error terms and for the regression of the smaller error terms, respectively.

The degrees of freedom for both the numerator and the denominator were calculated as follows:

$$\frac{n - k - 2m - 2}{2} = \frac{100 - 20 - 2(2) - 2}{2} = 37 \quad (5.97)$$

where $n$ is the total number of samples;

$k$ is the number of samples not included in either regression;

$m$ is the number of RHS variables in the equation, (does not include the intercept).

The F value for a significance level of 0.05 and for 37 degrees of freedom for both the numerator and the denominator is 1.69 (actually for degrees of freedom equal to 40), and so the height model was considered to be homoskedastic (identically distributed).

The estimated error versus predicted height ratio graph indicated that the selected model was somewhat in error; the graph of the weighted error terms versus the weighted
predicted height showed a somewhat linear trend for a few of the sample trees with small height ratios. The heteroskedasticity graph of estimated error squared with predicted height ratio showed that higher squared errors were associated with small predicted values; after examination, this trend was considered to be associated with lack-of-fit of the model rather than with heteroskedasticity. The Goldfeld and Quandt (1965) test was attempted and later discarded as the test is affected by lack-of-fit. The error terms for the height equation were therefore considered to be homoskedastic for the remaining analysis.

The estimated error versus predicted total volume graph indicated some heteroskedasticity, with larger variances for larger predicted values. The $R^2$ value for the simple linear regression of estimated error squared with predicted volume was 0.34276. The Goldfeld and Quandt (1965) test statistic was calculated and found to be $0.00495/0.00015$ or 33.000. The degrees of freedom were 38, as $m$ equals 1. Using the same F value as for the height model with degrees of freedom of 40 for the numerator and the denominator, the total volume model was found to be heteroskedastic.

The heteroskedasticity graph for the volume ratio model did not indicate heteroskedasticity. Also, a simple linear regression of the estimated error squared with the predicted volume ratio had an $R^2$ value of less than 0.005, and so the error terms were considered to be homoskedastic.

5.4.3 Estimating the Error Covariance Matrix of Each Equation

The error terms for all models were considered to be iid, except for the total volume equation which were found to be heteroskedastic. Several assumptions about the nature of the heteroskedasticity were examined.
Chapter 5. Application 1: Tree Volume Estimation

1. The variance of the error is a linear function of the predicted total volume.

\[ u^2_{2m} = \hat{e}^2_{2m} = \alpha_0 + \alpha_1 \text{ pred. total volume} + \text{error} \]  

where \( u^2_{2m} = \hat{e}^2_{2m} \) is the estimated error squared for the total volume equation (the third in the system) for the \( m^{th} \) sample, based on the unweighted model;

\( \text{pred. total volume} \) is based on the unweighted model;

\( \alpha_0, \alpha_1 \) are coefficients to be estimated.

2. The variance of the error is proportional to a power of the predicted total volume.

\[ u^2_{2m} = \hat{e}^2_{2m} = \alpha_3 (\text{pred. total volume})^{\alpha_4} \text{error} \]  

where \( \alpha_3 \), and \( \alpha_4 \) are coefficients to be estimated.

3. The variance of the error is a linear function of exogenous variables in the system.

\[ u^2_{2m} = \hat{e}^2_{2m} = f(X) \]  

where \( X \) is an \( n \) by \( k \) matrix of all of the exogenous variables in the system.

Each of these assumptions was assessed by using multiple linear regression to calculate coefficients. The estimated \( \alpha \) coefficients were then used to calculate the inverse of the square root of the estimated \( u^2_{2m} \), and these values were used to in a weighted regression of the total volume equation. The residual graph of the estimated error terms and predicted values of the weighted model were examined for a trend showing heteroskedasticity. None of these assumptions for heteroskedasticity resulted in a residual graph which indicated that the weighted error terms were identically distributed. Also, many negative variances were predicted using these equations for heteroskedasticity. As an alternative to these heteroskedasticity models, then, the variance of the error terms was calculated by dividing the range of the total volume into classes, and then the average squared error was
calculated for each of these classes. The residual graph of the weighted error terms with weighted predicted total volume, using the inverse of the square root of these estimated variances did not appear to be heteroskedastic. Also, the regression of the estimated weighted error terms squared versus the predicted weighted total volume resulted in an $R^2$ value of 0.05733, much lower than the $R^2$ value of 0.34276 for the unweighted error terms. The estimated variances of the error calculated for each total volume class were therefore used to weight the total volume regression.

5.4.4 Appropriate OLS Fit Based on Error Structure

The coefficients presented for the unweighted fit of the height, height ratio, and volume ratio models were considered appropriate as each of these equations have error terms which are iid. For the volume equation, EGLS using the estimated error covariance matrix from the previous section was used as an appropriate OLS fit. The estimated coefficients using the appropriate OLS fit for the system of equations were as follows:

\[
\begin{align*}
\text{pred. height} & = 0.584579 + 1.071239 \text{dbh} - 0.009644 \text{dbh}^2 \\
\text{pred. HR} & = 1.000000 - 0.912563 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 - 0.226424 hs \\
\text{pred. total volume} & = 0.011615 + 3.552805 \times 10^{-5} \text{dbh}^2 \text{height} \\
\text{pred. VR} & = 0.990691 - 0.986292 (1 - HR)^2
\end{align*}
\]

These fitted equations were used to compare the OLS method to other methods.

5.5 Composite Model Fit

5.5.1 Derivation of the Composite Model

To obtain a composite model for comparison to the other methods for fitting equations, the equations of the system were combined into one equation to estimate merchantable
volume by performing the following steps.

1. Height was replaced in the total volume equation by the height estimation function, resulting in the following equation.

$$\text{total volume} = \tau_0 + \tau_1 \text{dbh}^2 + \tau_2 \text{dbh}^3 + \tau_3 \text{dbh}^4 + \text{error}_{18} \quad (5.105)$$

This equation is simply a local volume function which is a polynomial.

2. The squared term of the volume ratio equation was expanded, and the height ratio was replaced by the height ratio equation.

$$VR = \tau_4 + \tau_5 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 + \tau_6 \text{hs} + \tau_7 \left( \frac{\text{top dib}}{\text{dbh}} \right)^4 + \tau_8 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 \text{hs}$$

$$+ \tau_9 \text{hs}^2 + \text{error}_{19} \quad (5.106)$$

3. The expanded total volume equation and the expanded volume ratio equation were then multiplied to obtain an equation to estimate merchantable volume.

$$\text{merch. volume} = \gamma_0 + \gamma_1 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 + \gamma_2 \text{hs} + \gamma_3 \left( \frac{\text{top dib}}{\text{dbh}} \right)^4 + \gamma_4 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 \text{hs}$$

$$+ \gamma_5 \text{hs}^2 + \gamma_6 \text{dbh}^2 + \gamma_7 \text{top dib}^2 + \gamma_8 \text{dbh}^2 \text{hs} + \gamma_9 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 \text{top dib}^2 + \gamma_{10} \text{top dib}^2 \text{hs}$$

$$+ \gamma_{11} \text{dbh}^2 \text{hs} + \gamma_{12} \text{dbh}^3 + \gamma_{13} \text{top dib}^2 \text{dbh} + \gamma_{14} \text{dbh}^3 \text{hs} + \gamma_{15} \frac{\text{top dib}^4}{\text{dbh}}$$

$$+ \gamma_{16} \text{top dib}^2 \text{dbh} \text{hs} + \gamma_{17} \text{dbh}^3 \text{hs}^2 + \gamma_{18} \text{dbh}^4 + \gamma_{19} \text{dbh}^2 \text{top dib}^2 + \gamma_{20} \text{dbh}^4 \text{hs}$$

$$+ \gamma_{21} \text{top dib}^4 + \gamma_{22} \text{dbh}^2 \text{top dib}^2 \text{hs} + \gamma_{23} \text{dbh}^4 \text{hs}^2 + \epsilon_5 \quad (5.107)$$

5.5.2 Unweighted Regression of Composite Model

The multiple linear regression of this large model resulted in the following fitted model.

$$\text{pred merch. volume} = 0.046284 + 2.69205 \times 10^{-5} \text{dbh}^3$$

$$-3.19287 \times 10^{-7} \text{dbh}^4 \text{hs} - 0.095285 \left( \frac{\text{top dib}}{\text{dbh}} \right)^4 \quad (5.108)$$
The addition of other variables produced an $R^2$ change of less than 0.001. The $R^2$ value for this composite model was 0.9682, with the first variable, $dbh^3$, responsible for a 0.9572 $R^2$ change. Because all of the RHS variables are exogenous, the estimated coefficients are unbiased.

5.5.3 Testing for iid Error Terms

Data were ordered by age measured on each tree at stump height and a graph of the estimated error term with the previous estimated error term was obtained. The graph did not indicate that the serial correlation was significant, and also, the linear regression of the estimated error term with the previous error term resulted in an $R^2$ value of 0.00508. The Durbin and Watson (1951) test statistic was also not significant. The error terms for the composite model were therefore independent.

A graph of the estimated error terms with the predicted merchantable volume from the unweighted linear model indicated that the variance of the error was heteroskedastic. Also, the linear regression of the estimated error terms squared with the predicted merchantable volume had an $R^2$ value of 0.13863. The test statistic for the Goldfeld and Quandt (1965) test was 16.4415, and since the critical value from the F distribution for $\alpha$ equal to 0.05, and 36 degrees of freedom (i.e. for $n = 100, k = 20, m = 3$) for the numerator and the denominator is 1.69 (actually for 40 degrees of freedom), the composite model was considered heteroskedastic.

5.5.4 Estimating the Variance of the Error Terms

The following model was selected for estimating the variance of the error terms of the composite model.

$$ u_m^2 = \epsilon_m^2 = \alpha_5 + \alpha_6 \text{ pred. merch. volume} + \text{error}_{21} \quad (5.109) $$
where $u_m^2 = \hat{e}_m^2$ is the estimated error term for the $m^{th}$ sample based on
the unweighted fit of the composite model;
$u_m^2$ estimates $\sigma_m^2$, the variance of the error term for the $m^{th}$ sample;
$\text{pred. merch. volume}$ is predicted merchantable volume from the unweighted
fit of the composite model;
$\alpha_5$ and $\alpha_6$ are coefficients to be estimated.

The fitted equation was as follows:

$$\hat{\sigma}_m^2 = \hat{u}_m^2 = -0.00023 + 0.01376 \text{ pred. merch. volume} \quad (5.110)$$

For a predicted merchantable volume of 0.01671 or less, the estimated variance is negative,
and so the estimated variance of the error term was reset to $0.672 \times 10^{-7}$, the value of
this equation for predicted merchantable volume of 0.01672. A graph of the estimated
weighted error versus the predicted weighted merchantable volume indicated that the
variances of the weighted error terms were homogenous. Also, the linear regression of
the estimated weighted error squared with the predicted weighted merchantable volume
resulted in an $R^2$ value of 0.003. The model chosen to estimate the variance of the error
terms was therefore considered appropriate.

### 5.5.5 Weighted Regression of the Composite Model

The estimated coefficients from weighted regression were as follows:

$$\text{pred. merch. volume} = -0.053441 + 3.34751 \times 10^{-5} \text{ dbh}^3$$
$$-6.46642 \times 10^{-7} \text{ dbh}^4 \text{ h.s} + 0.014345 \left(\frac{\text{top_dib}}{\text{dbh}}\right)^4 \quad (5.111)$$

The results from this weighted fit were used to compare with other methods.
5.6 MSLS Fit

5.6.1 First Stage Equations

The chosen system of equations includes four equations which require estimation of the coefficients. Of these four equations, two equations, total volume and merchantable ratio, have endogenous variables appearing on the RHS. The total volume function is based on \( \text{dbh}^2 \times \text{height} \), and because \( \text{height} \) is an endogenous variable, this RHS variable is also endogenous. Similarly, the \((1 - HR)^2\) variable which appears on the RHS of the volume ratio equation is endogenous as \( HR \) is endogenous. The OLS fit of each of these two equations therefore results in biased estimates of the coefficients. The error terms are likely correlated among equations, because a measure of taper over the stem was not included in the equations. Taper likely affects all four of the equations, resulting in correlation of the error terms. The system of equations met the rank and order conditions for identification.

The first step to obtain the MSLS fit was to fit the endogenous variables which appear on the RHS, \( \text{height} \) and \( HR \), as a function of all of the exogenous variables in the system of equations. Multiple linear regression was used to obtain the following first stage equations.

\[
\text{pred. height}_{1st} = 0.623596 + 1.092101 \text{dbh} - 0.009839 \text{dbh}^2 + 0.598333 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 - 2.568663 \text{hs} \tag{5.112}
\]

\[
\text{pred. HR}_{1st} = 0.979888 + 0.001844 \text{dbh} - 3.77511 \times 10^{-5} \text{dbh}^2 - 0.911111 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 + 0.227930 \text{hs} \tag{5.113}
\]

Some loss in efficiency is expected, because the \( HR \) variable appears as a quadratic term on the RHS of the system (nonlinear variable) and as a linear variable on the LHS of the system; the structural equations cannot be manipulated to obtain linear reduced-form
equations. Since the first stage equations predict the linear term, some efficiency was lost. The MSLS fit using these first stage equations will not, therefore, be asymptotically efficient.

### 5.6.2 Second Stage Equations

The second step of the MSLS procedure was to replace the height and HR variables appearing on the RHS of the volume and volume ratio equations by \( \text{pred. height}_{1st} \) and \( \text{pred. HR}_{1st} \). Each second stage equation was then fitted using multiple linear regression. Because the height and height ratio equations do not have endogenous variables on the RHS, the final OLS results were used as the second stage regressions of these two models. For the volume and volume ratio models, the following second stage equations were obtained.

\[
\text{pred. total volume}_{2nd} = 0.009677 + 3.57295 \times 10^{-5} \text{dbh}^2 \text{pred. height}_{1st} \quad (5.114)
\]

\[
\text{pred. } VR_{2nd} = 0.982007 - 0.934982 (1 - \text{pred. HR}_{1st})^2 \quad (5.115)
\]

### 5.6.3 Testing for iid Error Terms

The height and height ratio second stage equations are the same as the OLS equations and these equations were shown to have iid error terms. For the total volume and volume ratio equations, because the equations were purged of simultaneity bias, the coefficients of the second stage total volume and volume ratio equations are consistent, and so the residuals can be used as estimates of the error terms.

To test each equation for serial correlation, the estimated error terms for these two second stage equations were first ordered by the age of tree taken at stump height, and graphs of the estimated error term versus the previous estimated error term were obtained. Neither equation appeared to be serially correlated and the Durbin and Watson
(1951) test statistic was not significant for an alpha level of 0.05 for positive serial correlation and for an alpha level of 0.05 for negative correlation (alpha of 0.10 for a two sided test). Errors terms for both equations were therefore considered independent.

To check for heteroskedasticity in the second stage total volume model, a graph of the estimated error versus the predicted total volume was examined. The graph indicated increasing variance of the error term with increasing predicted total volume and the $R^2$ value for the linear regression of the estimated error squared with the predicted total volume was 0.20678. The test statistic for the Goldfeld and Quandt (1965) test was 22.11 which is significant for alpha equal to 0.05. The error terms for the second stage total volume equation were therefore considered heteroskedastic.

The graph of the estimated error with the predicted volume ratio indicated a lack-of-fit for a few samples with small volume ratios. The Goldfeld and Quandt (1965) test was calculated, but later discounted as the test is inconclusive if the model indicates lack-of-fit. No further testing or model development was done as this lack-of-fit is likely due to the height ratio model, and a more intensive examination of the height ratio model was considered to be beyond the scope of this research.

5.6.4 Estimation of the Error Covariance Matrix

All of the equations of the system which require estimation of coefficients have iid error terms except the total volume equation. The estimation of the error covariance matrix, then, first required the estimation of the variances of the error terms for the total volume equation. However, as with the estimation of the variances of the error terms for the total volume equation using the OLS method, none of the models proposed for heteroskedasticity resulted in weighted error terms which were iid. For this reason, and also to maintain a parallel fit for comparison to the OLS fitting technique, the range of the total volume was divided into classes, and the average squared estimated error was
calculated for each class. These averages were used as the estimated variances for the error terms.

Secondly, to estimate the error covariance matrix, EGLS was used to obtain estimates of coefficients for the weighted total volume model. Contemporaneous variances were then calculated using the estimated error terms from the following equations.

\[
\text{pred. height}_{2nd} = 0.584579 + 1.071239 \text{dbh} - 0.009644 \text{dbh}^2 
\]

\[
\text{pred. HR}_{2nd} = 1.000000 - 0.912563 \left( \frac{\text{top dib}}{\text{dbh}} \right)^2 - 0.226424 \text{hs}
\]

\[
\text{pred. total volume}_{2nd, wted} = 0.004272 \text{wt} + 3.590292 \times 10^{-5} \text{dbh}^2 \text{pred. height}_{1st, wted}
\]

\[
\text{pred. VR}_{2nd} = 0.982007 - 0.934982 (1 - \text{pred. HR}_{1st})^2
\]

where \( \text{wt} \) is the inverse of the square root of the estimated variance of the error term; 

\( \text{pred. total volume}_{2nd, wted} \) is total volume times the weight; 

\( \text{dbh}^2 \text{pred. height}_{1st, wted} \) is dbh squared times predicted height (from first stage equations) times the weight.

Contemporaneous variances were calculated using equation 3.28, resulting in the following \( \hat{\Sigma} \) matrix.

\[
\hat{\Sigma} = \begin{bmatrix}
6.6046 & -0.0022 & 2.0391 & -0.0089 \\
-0.0022 & 0.0019 & 0.0098 & 0.0019 \\
2.0391 & 0.0098 & 0.9606 & 0.0028 \\
-0.0089 & 0.0019 & 0.0028 & 0.0034
\end{bmatrix}
\]

where diagonal elements are the variances for an equation and off-diagonal elements are covariances between equations.

Based on equation 3.51, the estimated error covariance matrix was therefore the following
matrix.

\[
\hat{\Omega} = \begin{bmatrix}
6.6046W_1W_1' -0.0022W_1W_2' & 2.0391W_1W_3' & -0.0089W_1W_4' \\
-0.0022W_2W_1' & 0.0019W_2W_2' & 0.0098W_2W_3' & 0.0019W_2W_4' \\
2.0391W_3W_1' & 0.0098W_3W_2' & 0.9606W_3W_3' & 0.0028W_3W_4' \\
-0.0089W_4W_1' & 0.0019W_4W_2' & 0.0028W_4W_3' & 0.0034W_4W_4'
\end{bmatrix}
\] (5.121)

Because the height, height ratio, and volume ratio equations have iid error terms, \( W_1, W_2, \) and \( W_4 \) are equal to the identity matrix of size \( n \) by \( n \). The \( W_3 \) matrix is an \( n \) by \( n \) matrix with the diagonal elements equal to the square root of the estimated variances of the error terms for the second stage total volume model. The simplified error covariance matrix was therefore as follows:

\[
\hat{\Omega} = \begin{bmatrix}
6.6046I_n & -0.0022I_n & 2.0391W_3 & -0.0089I_n \\
-0.0022I_n & 0.0019I_n & 0.0098W_3 & 0.0019I_n \\
2.0391W_3 & 0.0098W_3 & 0.9606W_3W_3' & 0.0028W_3 \\
-0.0089I_n & 0.0019I_n & 0.0028W_3 & 0.0034I_n
\end{bmatrix}
\] (5.122)

The negative contemporaneous variances (covariances between equations) indicate that the height equation was negatively correlated with the height ratio and the volume ratio equations. All other equations appear positively correlated. These estimated covariances appear large relative to the estimated variances.

### 5.6.5 EGLS to Fit the System of Equations

To obtain the MSLS fit, EGLS was then used to fit the system of equations simultaneously for the estimated error covariance matrix. The resulting MSLS fit was as follows:

\[
pred.\ height_{MSLS} = -0.054629 + 1.100452dbh \\
-0.009568dbh^2
\] (5.123)
Chapter 5. Application 1: Tree Volume Estimation

\[
pred. HR_{MSLS} = 0.970531 - 0.898900 \left( \frac{\text{top dib}}{dbh} \right)^2 - 0.217870 \text{hs} \tag{5.124}
\]

\[
pred. \text{total volume}_{MSLS} = 0.003757 + 3.738763 \times 10^{-5} \ dbh^2 \ pred. \ height_{1st} \tag{5.125}
\]

\[
pred. VR_{MSLS} = 0.985298 - 0.947024 \ (1 - \ pred. \ HR_{1st})^2 \tag{5.126}
\]

The IMSLS technique was not used, because the number of samples was considered large enough to assume that the asymptotic properties of MSLS apply.

5.7 Comparison of the Three Fitting Techniques

5.7.1 Goodness-of-fit Measures

Fit Index

The Fit Indices for each of the four equations for the OLS and the MSLS fits are presented in Table 5.4. The Fit Indices for the MSLS fit were marginally lower than those for the OLS fit.

Table 5.4: Fit Indices for OLS and MSLS Fits of the Volume Equation System

<table>
<thead>
<tr>
<th>LHS Variable</th>
<th>height</th>
<th>HR</th>
<th>total volume</th>
<th>VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.7699</td>
<td>0.9722</td>
<td>0.9881</td>
<td>0.9960</td>
</tr>
<tr>
<td>MSLS</td>
<td>0.7612</td>
<td>0.9629</td>
<td>0.9629</td>
<td>0.9314</td>
</tr>
</tbody>
</table>

For the composite model, the Fit Index calculated for the merchantable volume was 0.9471. The corresponding value for the OLS fit, calculated by combining the total volume and volume ratio equations was 0.9885. Similarly, for the MSLS fit, the Fit
Index for merchantable volume was 0.9629. The composite model created by combining equations therefore resulted in the lowest Fit Index; the OLS Fit Index was the highest, but was only marginally higher than that for the MSLS fit.

**Mean Absolute Deviation**

The mean absolute deviations (M.A.D.) were calculated by class for each of the LHS endogenous variables. The classes were created by sorting the 100 samples by the endogenous variable, and then dividing the sorted data into five equal classes of 20 samples each. The M.A.D. values for the OLS fit and for the MSLS fit are presented in Table 5.5 and Table 5.6. The M.A.D. values are slightly lower for OLS than for MSLS. The trends across the five classes were basically the same for the OLS and for the MSLS fits.

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td></td>
<td>2.19</td>
<td>1.52</td>
<td>2.22</td>
<td>2.05</td>
<td>2.20</td>
</tr>
<tr>
<td>HR</td>
<td></td>
<td>0.053</td>
<td>0.036</td>
<td>0.026</td>
<td>0.016</td>
<td>0.0</td>
</tr>
<tr>
<td>total volume</td>
<td></td>
<td>0.009</td>
<td>0.013</td>
<td>0.026</td>
<td>0.027</td>
<td>0.064</td>
</tr>
<tr>
<td>VR</td>
<td></td>
<td>0.012</td>
<td>0.013</td>
<td>0.013</td>
<td>0.010</td>
<td>0.009</td>
</tr>
</tbody>
</table>

For the composite model, the M.A.D. values were calculated for merchantable volume, and compared to those for the OLS and MSLS fits, created by combining the total volume and volume ratio equations (Table 5.7). As with Fit Indices, the OLS fit gave the lowest M.A.D. values, followed by the MSLS fit, and lastly by the composite model fit.
Table 5.6: M.A.D. for Five Classes for the MSLS Fit of the Volume Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>height</td>
<td>2.00</td>
</tr>
<tr>
<td>HR</td>
<td>0.047</td>
</tr>
<tr>
<td>total volume</td>
<td>0.011</td>
</tr>
<tr>
<td>VR</td>
<td>0.078</td>
</tr>
</tbody>
</table>

Table 5.7: M.A.D. for Merchantable Volume

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>OLS</td>
<td>0.005</td>
</tr>
<tr>
<td>Composite</td>
<td>0.039</td>
</tr>
<tr>
<td>MSLS</td>
<td>0.013</td>
</tr>
</tbody>
</table>
Mean Deviation

The mean deviations (M.D.) were calculated for each equation for the same five classes as for M.A.D. Results for the OLS fit are presented in Table 5.8, and for the MSLS fit in Table 5.9. The M.D. values were sometimes lower for MSLS than for OLS. For the

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>height</td>
<td>-1.41</td>
</tr>
<tr>
<td>HR</td>
<td>-0.013</td>
</tr>
<tr>
<td>total volume</td>
<td>-0.008</td>
</tr>
<tr>
<td>VR</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 5.9: M.D. for Five Classes for the MSLS Fit of the Volume Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>height</td>
<td>-0.683</td>
</tr>
<tr>
<td>HR</td>
<td>0.005</td>
</tr>
<tr>
<td>total volume</td>
<td>-0.009</td>
</tr>
<tr>
<td>VR</td>
<td>-0.021</td>
</tr>
</tbody>
</table>

height model, the M.D. values were lower for smaller height values using MSLS than using OLS, but for higher height values, the OLS fit resulted in lower M.D. values. The MSLS fit of the height ratio model resulted in an underestimation of the height ratio for
all classes of the sample data; whereas, the OLS fit resulted in an overestimation of the lower classes and an underestimation of the higher classes. The M.D. values for total volume showed an overestimation of total volume using MSLS. The OLS fit showed small differences which were not consistently low or high across the classes. Finally, for the volume ratio model, the MSLS fit showed an overestimation for the lower classes, and an underestimation for the upper classes. The OLS fit showed underestimation, then overestimation and then underestimation.

For the composite model, the M.D. values were calculated for merchantable volume, and compared to those for the OLS and MSLS fits, created by combining the total volume and volume ratio equations (Table 5.10). The composite model underestimated merchantable volume for both smaller and larger volumes. The OLS and MSLS fits had similar trends with an overestimate of volume, then an underestimate, and finally an overestimate. The OLS values were the lowest.

5.7.2 Relative Variances

The trace of the estimated coefficient covariance matrix for the MSLS fit was 1.775283; whereas, the trace of the OLS fit was 3.467678 for the system of equations. Even though the MSLS appears more efficient, the trace of the estimated coefficient covariance matrix for the OLS fit cannot be used to calculate relative efficiency. The MSLS fit is expected
to be more efficient and in fact reaches the Cramer-Rao lower bound asymptotically, but in this case, because of the first stage equations used in this MSLS fit of the volume equation system, there was some loss in efficiency.

The traces for the submatrices of the estimated coefficient covariance matrix corresponding to each individual equation are given in Table 5.11. The trace for the OLS fit of the height equation was larger than the MSLS fit, and this difference accounted for the difference in the overall traces of coefficient covariance matrix for the system of equations. The lower trace for the MSLS fit occurred, because contemporaneous correlation is accounted for in the MSLS fit. A similar result occurred for the height ratio equation. For the total volume and volume ratio equations, the trace for the OLS fit was somewhat lower. However, since the OLS estimates of the coefficients are biased and inconsistent, confidence intervals cannot be calculated, and hypotheses cannot be tested.

Since the composite model has merchantable volume as the LHS variable and a different functional form, a comparison of the trace of the estimated coefficient covariance matrix for the composite model to the other fitting techniques was not possible.

### Table 5.11: Trace of the Coefficient Covariance Matrix for Each Equation of the Volume Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>height</th>
<th>HR</th>
<th>total volume</th>
<th>VR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>3.465530</td>
<td>2.097202 x 10^{-3}</td>
<td>9.954486 x 10^{-6}</td>
<td>4.125564 x 10^{-5}</td>
</tr>
<tr>
<td>MSLS</td>
<td>1.773650</td>
<td>9.787514 x 10^{-4}</td>
<td>1.483118 x 10^{-5}</td>
<td>6.084494 x 10^{-4}</td>
</tr>
</tbody>
</table>

5.7.3 Table of Estimated Coefficients and Standard Deviations

A summary of the estimated coefficients and their associated standard deviations from the OLS and MSLS fits is shown in Table 5.12.
Table 5.12: Estimated Coefficients and Standard Deviations of Coefficients for the Volume Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Coefficients</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>MSLS</td>
</tr>
<tr>
<td>height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{11}$</td>
<td>0.584579</td>
<td>-0.054629</td>
</tr>
<tr>
<td>$\delta_{12}$</td>
<td>1.071239</td>
<td>1.100452</td>
</tr>
<tr>
<td>$\delta_{13}$</td>
<td>-0.009644</td>
<td>-0.009568</td>
</tr>
<tr>
<td>HR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>1.000000</td>
<td>0.970531</td>
</tr>
<tr>
<td>$\delta_{22}$</td>
<td>-0.912563</td>
<td>-0.898900</td>
</tr>
<tr>
<td>$\delta_{23}$</td>
<td>-0.226424</td>
<td>-0.217870</td>
</tr>
<tr>
<td>total volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{31}$</td>
<td>0.0116150</td>
<td>0.003757</td>
</tr>
<tr>
<td>$\delta_{32}$</td>
<td>$3.552805 \times 10^{-5}$</td>
<td>$3.738763 \times 10^{-5}$</td>
</tr>
<tr>
<td>VR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{41}$</td>
<td>0.990691</td>
<td>0.985298</td>
</tr>
<tr>
<td>$\delta_{42}$</td>
<td>-0.986292</td>
<td>-0.947024</td>
</tr>
</tbody>
</table>
The estimated coefficients for the MSLS and for the OLS fits were very similar, except for the intercepts of the height and volume equations, and the slope of the volume ratio equation.

The standard deviations for the estimated coefficients of the height model were lower from the MSLS fit than from the OLS fit. This was reflected in the trace calculated for this equation, and indicates that the inclusion of the contemporaneous variances in the system fit using MSLS resulted in a lower variance. Results were similar for the fits of the height ratio equation. For the volume and volume ratio equations, the standard deviations for the coefficients were higher for the MSLS fit; again, this was reflected in the trace values for these equations, and indicates that OLS underestimates the confidence interval for the true value, because of simultaneity bias.

If the OLS estimates for the volume and volume ratio equations were used to test hypothesis statements, results would not be correct. For instance, $\delta_{41}$ would be expected to be equal to 1.0, so that a volume ratio of one is predicted, if the merchantable length is equal to the total height. Using an alpha level of 0.01, the confidence limits using the biased estimates from OLS would be 0.94923 to 0.98646. Using the consistent MSLS estimates, the confidence limits would be 0.96830 to 1.00230. The hypothesis that this coefficient is equal to 1.0 would be rejected with the OLS estimates, but not rejected with the MSLS consistent estimates.

5.7.4 Ranking for Other Features

The ranks assigned to each fitting technique are shown in Table 5.13. The composite model was assigned a rank of one for information as only one endogenous variable is estimated. For consistency, both the MSLS and the composite model for this system of equations provided consistent estimates. The fit of the composite model was unbiased.

---

4See Chapter 4 for an explanation of the ranking system used.
Table 5.13: Ranks for the Three Techniques

<table>
<thead>
<tr>
<th>Feature</th>
<th>OLS</th>
<th>Composite</th>
<th>MSLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Consistency</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Confidence Limits</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Asymptotic</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compatibility</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ease of fit</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

as all RHS variables were exogenous. Confidence limits for both the MSLS fit and the fit of the composite model can be calculated. MSLS was considered to have the highest asymptotic efficiency as the equations appear to be correlated. Both the MSLS and the composite model fits provide compatible estimates. The number of steps required to obtain the OLS fit was four for the unweighted fit, eight to check for iid errors, one to estimate the error covariance matrix, and one to obtain the weighted fit of the total volume model, for a total of 14 steps. For the composite model, three steps were required to derive the model, two steps to check for iid error terms, and three steps to obtain the unweighted fit, the estimated error covariance matrix, and the weighted fit, for a total of eight steps. For the MSLS fit, two steps were required to fit the first stage models, four steps for the second stage models (actually used OLS estimates for two of the second stage models), eight to check for iid errors (actually only four as results from OLS used for two of the models), one to estimate the weights for the volume equation, one to estimate the covariances among equations, and one to fit the system of equations, simultaneously, for a total of 17 steps.
5.8 Conclusion

The chosen system of equations to estimate tree volume was composed of seven equations. Four of these equations had coefficients which were to be estimated.

Using the OLS technique, two of the equations were found to have iid error terms, and were fitted using multiple linear regression. The error terms for the height ratio equation were considered iid, although a lack-of-fit was noted. The total volume equation had heteroskedastic error terms. Many models of this heteroskedasticity were tested and found to be inadequate. The variances of the error terms were therefore estimated by calculating the average of the estimated error terms, squared, from the unweighted fit, for classes of the predicted total volume values, also from the unweighted fit. The total volume equation was then refitted using EGLS.

A composite model was derived by combining the seven equations into one to estimate merchantable volume. The unweighted fit, using a criterion of a minimum $R^2$ change of 0.005, included only three of the 23 terms of the composite model. Error terms were found to be heteroskedastic and so the variances of the error terms were estimated. Multiple linear regression was then used to fit the weighted equation.

The first stage equations for the MSLS fit of the volume equation did not reflect the quadratic nature of the height ratio variable; MSLS estimates were therefore not asymptotically efficient for this system. The error terms for the second stage equations of the MSLS fit were found to be similar to the error terms from the unweighted OLS fit. However, a lack-of-fit was noted for the estimated error terms of the volume ratio equation; no attempt was made to explain the cause of this lack-of-fit. A second set of second stage equations were therefore derived in which the second stage total volume equation was weighted and fit using EGLS. MSLS was used to fit the system using the estimated error covariance matrix. IMSLS was not used to fit the equations, because the
number of samples was large.

The best goodness-of-fit measures (high F.I. and low M.A.D. and M.D. values) were obtained for the OLS fit. However, the goodness-of-fit measures for the MSLS fit were very similar. The composite model resulted in the worst goodness-of-fit measures, likely because some of the important variables were excluded from the composite model fit.

The trace of the estimated coefficient covariance matrix was lower for MSLS than for OLS. The trace values of the submatrices of this matrix, corresponding to individual equations, showed that the OLS fit resulted in lower trace values than the MSLS fit for the total volume and volume ratio equations. Each of these equations had RHS endogenous variables. The trace values for the OLS fit were higher for the height and height ratio equations than for the MSLS fit; these equations did not have RHS endogenous variables. The MSLS fit therefore resulted in an increase in efficiency as shown by the lower trace values for the submatrices estimated for the height and height ratio equations. Higher trace values for the submatrices of the total volume and volume ratio equations resulted from the MSLS fit. However, the OLS estimates of the variances cannot be used to calculate confidence intervals or test hypotheses concerning the true coefficients. The MSLS fit results in a consistent estimate of the coefficients and of the coefficient covariance matrix, and, therefore, confidence limits for true coefficients can be calculated. Hypotheses can also be tested using the MSLS fit.

The table of estimated coefficients and their standard deviations further supported the evidence given by the traces. The OLS fit results in larger standard deviations than the MSLS fit for the height and height ratio equations, and results in underestimates of confidence limits for the volume and volume ratio equations.

The summed rank for other features was the same for the MSLS fit as for the composite model fit. The MSLS technique has the advantage of all of the endogenous variables being predicted, whereas the composite model fit is easier to compute. The OLS fit
had the lowest summed rank, largely because estimates of the coefficients are biased and inconsistent. Also, the OLS fit is a single equation approach which does not result in compatible estimates, and efficiency is lost because the information concerning contemporaneous correlation is not utilized in the OLS fit.

The MSLS technique is therefore preferable for this application, because the goodness-of-fit measures are close to the OLS fit and higher than the composite model fit. Estimates of the coefficients and their standard deviations are consistent, and more efficient. This was indicated by a high summed rank for other features. The composite model fit resulted in unbiased estimates of the coefficients and their standard deviations; however, only one variable was estimated. In addition, the MSLS fit required only three more steps than the OLS fit.

The main disadvantage of the MSLS fit was that the number of samples had to be reduced from 500 to 100. This was due to a limitation of the computer memory to seven megabytes for this research. If a very large number of samples were used, and larger computer memory was not available, a subsample could be used, as was used for this application. Alternatively, the MSLS fit could be modified by eliminating the last step, which is EGLS applied to the system using the estimated error covariance matrix. Instead, consistent estimates of the coefficients could be obtained by using the weighted second stage coefficients (equations 5.116, 5.117, 5.118, and 5.119). The gain in efficiency by accounting for the contemporaneous variances in the final system fit would be lost, but simultaneity bias would be removed, which is an improvement over the OLS fit.
Chapter 6

Application 2: Estimation of Tree Diameter Distribution

6.1 Introduction

Diameter distribution information is required in order to choose stands for harvest and to assess the expected financial return. The estimation of diameter distributions from stand measures can be done by first selecting a known probability density function (pdf), finding the parameters, and then by relating the parameters of the selected pdf to current stand measures. These equations can then be combined with a stand level growth model where stand attributes are first predicted and parameters of the pdf are then “recovered” from these predicted stand attributes. Hyink and Moser (1983) referred to this diameter distribution modelling technique as the parameter recovery method. Stand attributes from forest inventories can also be used as inputs to predict the parameters of the selected pdf, once the relationships between each parameter and the stand attributes are established. The alternative to estimating the parameters of a known pdf is to relate the stand attributes (measured or predicted) to percentiles of the diameter distribution (Anon., 1987; Bailey et al., 1981; Borders et al., 1987b). To limit the scope of this thesis, the estimation of diameter distribution was restricted to the first method.

The most commonly selected pdf is the Weibull distribution (Clutter et al., 1983). The probability density function of this distribution is as follows:

\[
\begin{align*}
 f(X) &= \frac{c}{b} \left(\frac{X - a}{b}\right)^{c-1} e^{-\left(\frac{X - a}{b}\right)^{c}} \quad (a \leq X < \infty) \\
 f(X) &= 0 \quad \text{otherwise}
\end{align*}
\]  

(6.127)
where

\[ b \geq 0 \quad c \geq 0 \]

The parameter \( a \) is the location parameter, \( b \) is the scale parameter, and \( c \) is the shape parameter (Clutter et al., 1983). Although \( a \) is sometimes negative for the distribution, \( a \) must be nonnegative for diameter distributions. The Weibull distribution is attractive for representing diameter distributions in that the equation for the pdf is relatively simple with only three parameters, and the shape of the distribution is flexible (Bailey and Dell, 1973).

The cumulative form of the Weibull distribution, the cumulative density function (cdf), is as follows:

\[
F(X) = \begin{cases} 
1 - e^{-\left(\frac{X-a}{b}\right)^c} & (a \leq X < \infty) \\
0 & \text{otherwise}
\end{cases}
\]

(6.128)

A system of parameter prediction equations is used to predict the parameters of the Weibull distribution, which represents the distribution of diameters.

\[
\begin{align*}
a &= f_1(\text{stand attributes}) \\
b &= f_2(\text{stand attributes}) \\
c &= f_3(\text{stand attributes})
\end{align*}
\]

(6.129) (6.130) (6.131)

If the parameters of the Weibull distribution are considered interdependent, the system of parameter prediction equations becomes the following:

\[
\begin{align*}
a &= f_4(\text{stand attributes, } b, c) \\
b &= f_5(\text{stand attributes, } a, c) \\
c &= f_6(\text{stand attributes, } a, b)
\end{align*}
\]

(6.132) (6.133) (6.134)
A change in $b$, for example, is expected to change the values of $a$ and $c$ and these changes will only be reflected if all three parameters are included in each equation of the system. This second system is simultaneous in that the endogenous variables (the parameters) on the LHS also appear on the RHS.

The first step in estimating the diameter distribution using the Weibull pdf is to estimate the parameters of the pdf for each stand or plot. Estimates can be obtained by using nonlinear least squares to fit the cdf, and variances can be estimated by using Jackknife approximations or by using a Taylor series expansion. Variances of the parameters from this nonlinear fit of each plot are required and then the method of fitting systems of parameter prediction equations proposed by Ferguson and Leech (1978), modified for simultaneous systems of equations, could be followed. However, the error terms for each plot or stand may be heterogenous or serially correlated and a transformation or generalized nonlinear least squares would be required to obtain a consistent estimate of the coefficient covariance matrix. The alternative to nonlinear least squares is to replace the relative cumulative frequency observations in each plot or stand with three percentiles only and to calculate the $a$, $b$, and $c$ parameters from these three percentiles. The three parameters are considered to be without error with this method. This second approach was selected for this thesis; all system fitting problems were restricted to linear equations.

6.2 Preparation of Data

Summarized data for permanent sample plots (psp) were obtained from the Alberta Forest Service (AFS). Psp information was selected over temporary sample plot data because the psp data are fixed area plots, whereas most of the temporary sample plot data collected in Alberta are variable radius plots which are weighted toward selection of larger diameter trees. The summarized data included the location of the psp, the plot
size, the average age at breast height by species, the average age at stump height by
species, the site index for a reference age of 50 years at breast height by species, the
quadratic mean diameter (diameter of a tree with mean basal area) by species for live
trees, and the top height (height of the 100 largest trees by dbh per hectare) by species for
live trees. Also included were live stems per hectare by species and diameter class, and
volume from ground to tree tip (total tree volume) per hectare by species and diameter
class for live trees. The diameter classes had a width of 4.0 centimeters, beginning with
1.1 centimeters. A description of the data collection procedures currently used by the
AFS for psp data can be found in the *Alberta Forest Service: Permanent sample plot

Plots with more than 80 percent pine volume were selected from the data. Also, as
with the section tree data for the first application, data from pine stands in Northeastern
Alberta were removed and therefore only lodgepole pine plots remained. Prior to 1982,
a cluster of four plots was established by AFS at each psp location. To remove possible
dependencies of plots within these clusters, only one plot was retained from each location.
The psp data represented measurements from several time periods; to remove correla-
tion between measurements of a given plot over time, the first measurement alone was
retained. Also, plots which had been treated, for instance thinned plots, were removed
from the data.

The relative cumulative frequencies by diameter class (diameter distribution) for the
remaining plots were graphed. Plots which had multimodal diameter distributions were
removed, because the Weibull distribution is suitable only for unimodal distributions.
The 24th, 63rd, and 93rd percentiles were calculated for each psp, using linear interpo-
lation as shown below for the 24th percentile.

\[
\frac{\text{ratio} (k) - 0.24}{\text{ratio} (k) - \text{ratio} (k - 1)} = \frac{\text{endpoint} (k) - D1}{\text{endpoint} (k) - \text{endpoint} (k - 1)}
\]  

(6.135)
where \( ratio(k) \) is the ratio of the cumulative number of stems up to and including the \( k^{th} \) diameter class, over the total number of stems per hectare; \( k - 1 \) is the previous diameter class; the desired percentile lies between the two classes;

\( endpoint(k) \) and \( endpoint(k - 1) \) are the endpoints of the \( k^{th} \) and \( k - 1^{th} \) diameter classes, respectively;

\( D1 \) is the diameter corresponding to the 0.24 ratio.

Dubey (1967) showed that the 24\(^{th} \) and 93\(^{rd} \) percentiles are most efficient relative to the maximum likelihood method, if the \( a \) parameter is equal to zero. Bailey and others (1981) included the 63\(^{rd} \) percentile to estimate the Weibull distribution if \( a \) is unknown. The \( a \), \( b \) and \( c \) parameters of the Weibull distribution were then calculated from these percentiles by using the following equations.

\[
D1 = a + b \left[ -\ln \left(0.76\right) \right]^{1/c}
\]
\[
D2 = a + b \left[ -\ln \left(0.37\right) \right]^{1/c} \approx a + b
\]
\[
D3 = a + b \left[ -\ln \left(0.07\right) \right]^{1/c}
\]

where \( D1 \), \( D2 \), and \( D3 \) are the diameter limits for the 0.24, 0.63, and 0.93 cumulative probabilities, respectively;

\( \ln \) is the natural log, base \( e \).

These equations can be simplified to the following:

\[
\frac{D1 - D2}{D3 - D2} = \frac{\left[ -\ln \left(0.76\right) \right]^{1/c} - 1}{\left[ -\ln \left(0.07\right) \right]^{1/c} - 1}
\]
\[
b = \frac{D1 - D2}{\left[ -\ln \left(0.76\right) \right]^{1/c} - 1}
\]
\[
a = D2 - b
\]

An iterative procedure is required to solve for \( c \), and then \( b \) and \( a \) are calculated from the estimated \( c \) value. The Nonlinear Function Optimization (NLP) package at the
University of British Columbia (Vaessen, 1984) was used to solve for the $c$ parameter by minimizing the absolute difference between the right and left hand sides of equation 6.139. A tolerance limit of $1.0 \times 10^{-12}$ was used. For three of the unimodal psps, several different starting points were attempted, but no solution was found within the tolerance limit specified. In all of these three plots, the trees were in the two smallest diameter classes, only, resulting in an abrupt peak at the beginning of the graph of the distribution, followed by a sharp decline. The Weibull distribution is probably inappropriate for these plots, so the plots were deleted from further analysis. For two of the fitted plots, the solution for the $a$ parameter was negative. Since this negative value was close to zero, no attempt was made to refit these two plots.

The distribution of the remaining 121 psps by age and stems per hectare is presented in Table 6.14. Over half of the psps did not have age at breast height on file. These data were missing as the ages which had been previously recorded were considered to be in error by AFS personnel; new ages were taken, but these had not yet been added to the data used in this thesis. Site index was also unavailable for these plots. Data were checked for outliers by graphing measurements for pairs of variables. No outliers were found.

6.3 Model Selection

Multiple linear regression was used to fit each of the parameters as a function of the other parameters and stand measures. The stand measures initially included in the regression were site index for pine, age at breast height for pine, stems per hectare for all species, quadratic mean diameter for all species, top height for all species, total volume per hectare for all species. These variables were chosen based on the results in literature including a study by the AFS for predicting percentiles of diameter distributions for stands in Alberta.
Table 6.14: Distribution of Selected Psps by Age and Stems per Hectare Classes

<table>
<thead>
<tr>
<th>Stems per Hectare</th>
<th>Age at Breast Height</th>
<th>0 to 20</th>
<th>21 to 40</th>
<th>41 to 60</th>
<th>61 to 80</th>
<th>81 to 100</th>
<th>101 to 120</th>
<th>121 to 140</th>
<th>141 to 160</th>
<th>161 to 180</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 1000</td>
<td>0 to 1000</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>1001 to 2000</td>
<td>1001 to 2000</td>
<td>18</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>2001 to 3000</td>
<td>2001 to 3000</td>
<td>23</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>3001 to 4000</td>
<td>3001 to 4000</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>4001 to 5000</td>
<td>4001 to 5000</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>5001 to 6000</td>
<td>5001 to 6000</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>6001 to 7000</td>
<td>6001 to 7000</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7001 to 8000</td>
<td>7001 to 8000</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8001 to 9000</td>
<td>8001 to 9000</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9001 to 10000</td>
<td>9001 to 10000</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10000 to 11000</td>
<td>10000 to 11000</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>11000 to 12000</td>
<td>11000 to 12000</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>72</td>
<td>0</td>
<td>11</td>
<td>13</td>
<td>13</td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>121</td>
</tr>
</tbody>
</table>
(Anon., 1987). Only 49 of the 121 plots were used for this initial analysis, as age and site index were missing from the other 72 plots. Because the error terms were not assumed to be normally distributed or iid, the partial F test used in stepwise regression was not appropriate. Instead, all variables were forced into the regression. The addition of age and site index to the regression resulted in an \( R^2 \) change of less than 0.005; therefore, the multiple linear regression of all of the 121 plots was fitted, using all of the stand variables except age and site index. Based on a minimum of 0.005 change in \( R^2 \) value, the system of parameter prediction equations was selected as follows:

\[
a = \delta_{11} + \delta_{12} \text{topht} + \delta_{13} \text{qdiam} + \delta_{14} \text{stems} + \delta_{15} \text{totvol} + \delta_{16} b + \delta_{17} c + \epsilon_1 \tag{6.142}
\]

\[
b = \delta_{21} + \delta_{22} \text{topht} + \delta_{23} \text{qdiam} + \delta_{24} \text{stems} + \delta_{25} \text{totvol} + \delta_{26} a + \delta_{27} c + \epsilon_2 \tag{6.143}
\]

\[
c = \delta_{31} + \delta_{32} \text{topht} + \delta_{33} \text{qdiam} + \delta_{34} \text{stems} + \delta_{35} a + \delta_{36} b + \epsilon_3 \tag{6.144}
\]

where \text{topht} is the top height or the average height of the 100 largest trees by dbh per hectare;

\text{totvol} is the total volume from ground to tree tip per hectare;

\text{stems} is the number of stems per hectare;

\text{qdiam} is the mean dbh for the tree of average basal area;

\( \delta_{11} \) through \( \delta_{36} \) are coefficients to be estimated.

Since the equations for the \( a \) and \( b \) parameter are the same using this 0.005 \( R^2 \) criterion, these two equations cannot be distinguished from each other. A new criterion of a minimum \( R^2 \) change of 0.02 was set. The selected system of parameter prediction equations was therefore as follows:

\[
a = \delta_{11} + \delta_{12} \text{topht} + \delta_{13} \text{stems} + \delta_{14} \text{totvol} + \delta_{15} b + \epsilon_1 \tag{6.145}
\]

\[
b = \delta_{21} + \delta_{22} \text{qdiam} + \delta_{23} \text{stems} + \delta_{24} \text{totvol} + \delta_{25} a + \epsilon_2 \tag{6.146}
\]

\[
c = \delta_{31} + \delta_{32} \text{topht} + \delta_{33} a + \delta_{34} b + \epsilon_3 \tag{6.147}
\]
Chapter 6. Application 2: Estimation of Tree Diameter Distribution

Each of the equations is unique using this criterion for retaining variables. The system is simultaneous in that the endogenous variables, the parameters to be predicted, appear on both the RHS and the LHS of equations in the system. This second system of equations was used in the subsequent analyses.

6.4 Ordinary Least Squares Fit

6.4.1 Unweighted Simple or Multiple Linear Regression

The estimated coefficients for the chosen parameter prediction equations using unweighted multiple linear regression for each equation were as follows:

\[
pred. a = 7.214783 + 0.260097 \text{topht} - 9.558752 \times 10^{-4} \text{stems} \\
+ 0.026935 \text{totvol} - 1.225793 b \tag{6.148}
\]

\[
pred. b = 3.639838 + 0.159000 \text{qdiam} - 1.96650 \times 10^{-4} \text{stems} \\
+ 0.007499 \text{totvol} - 0.209573 a \tag{6.149}
\]

\[
pred. c = 4.436816 - 0.349270 \text{topht} + 0.237814 a + 0.540540 b \tag{6.150}
\]

The estimated error terms using these unweighted fitted equations were used to test for serial correlation and heteroskedasticity.

6.4.2 Testing for iid Error Terms

To check for serial correlation in each equation, the estimated error terms were sorted by the predicted LHS endogenous variable using the unweighted fit, and a graph of the current error term versus the previous error term was obtained. Simple linear regression was used to fit the regression of the current error with the previous error term for each of the three equations in the system. The \(R^2\) value for the equation with \(a\) as the LHS endogenous variable was 0.01937, for \(b\) the \(R^2\) was 0.00041, and for \(c\) the \(R^2\) was 0.01379.
The Durbin and Watson (1951) test statistic was not significant for any of the equations using an alpha level of 0.10 (0.05 for negative serial correlation and 0.05 for positive serial correlation). The error terms were therefore considered to be independent for each of the three equations.

A graph of the estimated error terms versus the predicted LHS endogenous variables was examined for heteroskedasticity. The simple linear regression of the estimated error squared with the predicted LHS endogenous variable resulted in an $R^2$ value of 0.0030 for the $a$ parameter, 0.1327 for the $b$ parameter, and 0.02366 for the $c$ parameter. The error terms for the first equation, with the $a$ parameter as the LHS endogenous variable, were considered to be homoskedastic. The Goldfeld and Quandt (1965) test was used to test for heteroskedasticity for the $b$ and $c$ parameter equations. Data were ordered by the predicted LHS endogenous variable, and linear regressions were performed for the first 40 and for the last 40 observations. The middle 41 observations were excluded from both regressions. The test statistic was found to be significant for each of these equations. The error terms for these two equations were therefore considered to be heteroskedastic.

### 6.4.3 Estimating the Error Covariance Matrix of Each Equation

The error terms for the first equation with parameter $a$ as the LHS endogenous variable were iid. For the remaining two equations of the system, the error terms were considered heterogenous.

For the $b$ parameter equation, the variances of the error terms were estimated using the following model.

$$ \hat{u}_{2m}^2 = 0.120250 \times \text{pred. } b \quad (6.151) $$

where $\hat{u}_{2m}^2$ is the estimated variance of the error for the second equation of the system and the $m^{th}$ sample. The $u_{2m}$ values are equal to the $\hat{e}_{2m}$ values from the
unweighted OLS fit;

\[ \text{pred. } b \] is the predicted \( b \) from the unweighted regression.

The \( \text{pred. } b \) variable was reset to \( 1.0 \times 10^{-7} \) if the value was less than or equal to zero resulting in only positive estimated variances. A linear model to estimate variances which included an intercept was also fitted, but the intercept was negative resulting in more of the estimated error terms as negative values. The model with an intercept was discarded and the zero intercept model was chosen. Using the inverse of the square root of these estimated variances as weights, a weighted regression of the \( b \) parameter equation was performed. The estimated error terms from this weighted model were graphed against the predicted weighted \( b \) parameter. This graph showed some heteroskedasticity, and the regression of the estimated error terms squared, from the weighted model, versus the predicted weighted \( b \) parameter resulted in an \( R^2 \) value of 0.09354. However, this \( R^2 \) value was somewhat lower than the \( R^2 \) value of 0.1327 for the unweighted model; these estimated variances were selected as weights for the \( b \) parameter equation.

The \( c \) parameter equation also appeared to have non-iid error terms, although the \( R^2 \) value for the regression of the estimated error terms squared with the predicted \( c \) values was only 0.02366. The average of the estimated error terms squared, for classes of the predicted \( c \) parameter indicated that the variance of the error terms was large for small values of the \( c \) parameter, small for medium values of the \( c \) parameter, and again large for large values of the \( c \) parameter. The Goldfeld and Quandt (1965) test for data ordered by the predicted \( c \) parameter was therefore considered to be misleading as the variances of the error did not increase with the predicted value. The average squared error terms by classes of the predicted \( c \) parameter were then used to obtain a weighted regression by using the inverse of the square root of the average estimated error squared as weights. The \( R^2 \) of the regression of the estimated weighted error terms, squared, with the predicted weighted \( c \) parameter was 0.05860. However, since the heteroskedasticity
was likely the result of a lack-of-fit, the error terms were considered iid for this analysis.

6.4.4 Appropriate OLS Fit Based on Error Structure

The coefficients presented for the unweighted fit of the \(a\) and \(c\) parameter equations were considered appropriate as error terms were iid. For the \(b\) parameter equation, EGLS was used to estimate the parameters, using the estimated variances of the error terms. The estimated coefficients using the appropriate OLS fit for the system of equations were as follows:

\[
\text{pred. } a = 7.214783 + 0.260097 \text{topht} - 9.558752 \times 10^{-4} \text{stems} + 0.026935 \text{totvol} - 1.225793 \times 10^{-4} \text{stems} - 1.225793 \times 10^{-4} \text{stems} \\
\text{pred. } b = 3.574736 + 0.162501 \text{qdiam} - 1.795836 \times 10^{-4} \text{stems} + 0.007432 \text{totvol} - 0.212128 \times 10^{-4} \text{stems} - 0.212128 \times 10^{-4} \text{stems} \\
\text{pred. } c = 4.436816 - 0.349270 \text{topht} + 0.237814 a + 0.540540 b + 0.540540 b
\]

These fitted equations were used to compare the OLS method to other methods.

6.5 Composite Model Fit

6.5.1 Derivation of the Composite Model

To obtain a composite model, the equations for the \(a\), \(b\) and \(c\) parameters were substituted into the cumulative form of the Weibull distribution function by performing the following steps.

1. The \(b\) parameter which appears on the RHS of the \(a\) parameter equation was replaced by the \(b\) parameter equation to obtain the following equation, after simplification.

\[
a = \alpha_{11} + \alpha_{12} \text{topht} + \alpha_{13} \text{qdiam} + \alpha_{14} \text{stems} + \alpha_{15} \text{totvol} + \text{error}_1
\]
2. The $a$ parameter which appears on the RHS of the $b$ parameter equation was replaced by the $a$ parameter equation to obtain the following equation.

$$ b = \alpha_{21} + \alpha_{22} \text{topht} + \alpha_{23} \text{diam} + \alpha_{24} \text{stems} + \alpha_{25} \text{totvol} + \text{error}_2 \quad (6.156) $$

3. The equations from above, for the $a$ and $b$ parameters as functions of stand variables only, were substituted for the $a$ and $b$ parameters of the $c$ parameter equation.

$$ c = \alpha_{31} + \alpha_{32} \text{topht} + \alpha_{33} \text{diam} + \alpha_{34} \text{stems} + \alpha_{35} \text{totvol} + \text{error}_3 \quad (6.157) $$

These substitutions resulted in equations to predict each of the parameters of the Weibull distribution as a function of the stand variables only. These three equations were then substituted into the cumulative form of the Weibull distribution to obtain the following equation.

$$ F(X) = 1 - e^{-\left(\frac{X}{b \text{ eqn}}\right)^{c \text{ eqn}}} + \epsilon_4 \quad (6.158) $$

where $a \text{ eqn}$, $b \text{ eqn}$, and $c \text{ eqn}$ are the equations given above, excluding the error terms.

### 6.5.2 Unweighted Regression of Composite Model

The composite equation was fitted using a nonlinear optimization routine (Vaessen, 1984) to minimize the sum of squared deviations between the actual and predicted values for the three percentiles of the 121 plots, simultaneously. Starting points for the nonlinear fit were the estimated coefficients obtained from the multiple linear regression of each of the three parameters as a function of the stand variables only. For each iteration, the estimated coefficients for the composite model were used to calculate estimated $a$, $b$, and $c$ parameters. Because the $b$ and $c$ parameters of the Weibull equation must be greater than zero, and the $a$ parameter must be nonnegative, the sum of squared deviations was calculated by taking the absolute value of each negative parameter. The fitted function
from the last iteration (minimum sum of squared error) resulted in seven of the 121 plots having an estimated $b$ parameter which was negative. Since the $b$ parameter is the denominator of a ratio and the numerator of this ratio was positive, the ratio was negative when $b$ was negative. Since this ratio is raised to the $c$ power, a negative $b$ parameter resulted in a mathematical error. To attempt to restrict the resulting fit for a nonnegative $a$ parameter, and $b$ and $c$ parameters greater than zero, one was added to the absolute value of negative parameters. The fitted function resulted in estimated $b$ and $c$ parameters which were greater than zero for all 121 plots. Some of the $a$ parameters were negative; since these values were close to zero and the $a$ parameter is the location parameter, no further attempts were made to restrict the fitted function.

The fitted composite model, partitioned into the $aeqn$, the $beqn$ and the $ceqn$ was as follows:

$$\begin{align*}
 pred. a eqn &= 5.051160 - 0.377366 \text{topht} + 0.002504 \text{qdiam} \\
 &\quad + 0.000263 \text{stems} + 0.092034 \text{totvol} \\
\text{pred. b eqn} &= 0.787755 + 0.559818 \text{topht} + 0.004303 \text{qdiam} \\
 &\quad - 0.000433 \text{stems} + 0.275654 \text{totvol} \\
\text{pred. c eqn} &= 7.468789 - 0.065044 \text{topht} + 0.014784 \text{qdiam} \\
 &\quad - 0.000122 \text{stems} - 0.214141 \text{totvol}
\end{align*}$$

(6.159) (6.160) (6.161)

6.5.3 Testing for iid Error Terms and Weighted Regression

Because the composite model is a nonlinear model, no testing of the error terms for serial correlation or heteroskedasticity was performed. The unweighted fit was therefore used to compare to the other two methods. Comparison was restricted to goodness-of-fit measures only.
Chapter 6. Application 2: Estimation of Tree Diameter Distribution

6.6 MSLS Fit

6.6.1 First Stage Equations

Each of the parameter prediction equations has endogenous variables on the RHS. The system of equations was identified.

The first step to obtaining an MSLS fit was to fit each of the RHS endogenous variables, parameters $a$ and $b$ as a function of all of the exogenous variables in the system. Multiple linear regression was used to obtain the following first stage equations.

\[
pred. a_{1st} = 6.014739 + 0.054645 \text{topht} - 0.066945 \text{qdiam} - 9.60807 \times 10^{-4} \text{stems} + 0.024632 \text{totvol}
\]
\[
pred. b_{1st} = 1.476369 + 0.057970 \text{topht} + 0.156755 \text{qdiam} + 4.26610 \times 10^{-5} \text{stems} + 0.001511 \text{totvol}
\]

A first stage equation was not required for parameter $c$, because $c$ appears only on the LHS side of the system of equations.

6.6.2 Second Stage Equations

The second stage equations were derived by replacing the endogenous variables, $a$ and $b$ which appear on the right hand side of equations in the parameter prediction system with the predicted endogenous variables from the first stage equations. Multiple linear regression was then used to fit each of the second stage equations.

\[
pred. a_{2nd} = 6.645248 + 0.079403 \text{topht} - 9.42588 \times 10^{-4} \text{stems}
\]
\[+0.025278 \text{totvol} - 0.427068 \text{pred. b}_{1st}
\]
\[
pred. b_{2nd} = -4.904405 + 0.227774 \text{qdiam} + 0.001062 \text{stems}
\]
\[-0.024620 \text{totvol} + 1.060856 \text{pred. a}_{1st}
\]
\[ \text{pred.} \, c_{2nd} = 5.177297 - 0.300508 \, \text{topht} + 0.331522 \, \text{pred.} \, a_{1st} + 0.064493 \, \text{pred.} \, b_{1st} \] (6.166)

where \( \text{pred.} \, a_{1st} \) and \( \text{pred.} \, b_{1st} \) are the predicted values for \( a \), and \( b \), respectively, using the first stage equations.

### 6.6.3 Testing for iid Error Terms

Each of the second stage equations was purged of the simultaneity bias by using predicted values from the first stage equations to replace RHS endogenous variables. The residuals from the second stage equations were therefore used as estimates of the error terms, because estimated coefficients are consistent.

To test for serially correlated error terms, the estimated error terms for each equation were ordered by the predicted LHS endogenous variable from the second stage equations, and graphs of the estimated error term versus the previous error term were examined. The \( R^2 \) value for the regression of the estimated error term with the previous error term was less than 0.005 for the \( b \) parameter equation. For the \( a \) parameter equation, the \( R^2 \) value was 0.04099, The Durbin and Watson (1951) test was inconclusive for positive serial correlation (alpha of 0.05), and was not significant for negative serial correlation (alpha of 0.05). The runs test was therefore used to test for serial correlation, positive or negative, using an alpha level of 0.10. The test statistic was not significant. For the \( c \) parameter equation, the \( R^2 \) value was 0.01215; the Durbin and Watson test was not significant for positive or negative serial correlation. The error terms were therefore considered to be independent for all three equations.

To check for heteroskedasticity, the regression of the estimated error squared with the predicted LHS endogenous variable was performed. The \( R^2 \) value for the \( a \) parameter equation was 0.0020, for the \( b \) parameter equation was 0.1137, and for the \( c \) parameter
equation was 0.0147. The $a$ parameter equation was therefore considered homoskedastic. The Goldfeld and Quandt (1965) test for the $b$ parameter equation was significant (test statistic of 3.04149) for an alpha level of 0.05, excluding the 41 observations representing middle values of the predicted $b$ parameter. The $b$ parameter was therefore considered to have heteroskedastic error terms. An examination of the estimated error terms for the $c$ parameter equation indicated that the variances of the error were likely not monotonic, as with the OLS fit, since this may be associated with a lack of fit, this equation was considered to be homoskedastic.

### 6.6.4 Estimation of the Error Covariance Matrix

The estimation of the error covariance matrix first required the estimation of the variances of the error terms for the $b$ parameter equation. The fitted equation for the estimated variances of the error term was as follows:

$$ \hat{u}_{2m}^2 = 0.164154 \times \text{pred. } b_{2nd} $$ (6.167)

where $\hat{u}_{2m}^2$ estimated variance of the error term for the second equation and the $m^{th}$ sample;

$\text{pred. } b_{2nd}$ is the predicted $b$ parameter from the second stage equations.

The predicted $b$ parameter was reset to $1.0 \times 10^{-7}$ if the value was zero or negative, so that only positive values were predicted for the variance of the error. The $R^2$ value for the regression of the estimated error terms, squared, with the predicted $b$ parameters from the weighted second stage model was 0.07497, which is somewhat less than the $R^2$ value of 0.1137 from the fit of the unweighted second stage model.

These estimated variances of the error terms were used to obtain an EGLS fit of the $b$ parameter equation. Contemporaneous variances were then calculated using the
estimated error terms from the following equations.

\[
\text{pred. } a_{2nd} = 6.645248 + 0.079403 \text{topht} - 9.42588 \times 10^{-4} \text{stems}
+ 0.025278 \text{totvol} - 0.427068 \text{pred. } b_{1st}
\]

(6.168)

\[
\text{pred. } b_{2nd wt} = -6.591194 wt + 0.242275 \text{q diam wt} + 0.001292 \text{stems wt}
- 0.030403 \text{totvol wt} + 1.302598 \text{pred. } a_{1st wt}
\]

(6.169)

\[
\text{pred. } c_{2nd} = 5.177297 - 0.300508 \text{topht} + 0.331522 \text{pred. } a_{1st}
+ 0.064493 \text{pred. } b_{1st}
\]

(6.170)

where \( wt \) is the inverse of the square root of the estimated variance of the error term;

\( \text{pred. } b_{2nd wt} \) is the \( b \) parameter times the weight;

\( \text{q diam wt} \) is quadratic mean diameter times the weight;

\( \text{stems wt} \) is stems per hectare times the weight;

\( \text{totvol wt} \) is volume per hectare times the weight;

\( \text{pred. } a_{1st wt} \) is the predicted \( a \) parameter times the weight.

Contemporaneous variances were calculated using equation 3.28, resulting in the following \( \hat{\Sigma} \) matrix.

\[
\hat{\Sigma} = \begin{bmatrix}
5.2812 & -1.1863 & 0.4059 \\
-1.1863 & 0.8688 & 0.2898 \\
0.4059 & 0.2898 & 2.8225
\end{bmatrix}
\]

(6.171)

where diagonal elements are the variances for an equation and off-diagonal elements are covariances between equations.

The error terms between the \( a \) and \( b \) parameter equations appeared negatively correlated, whereas the \( c \) parameter was positively correlated with the other two equations of the system. Based on equation 3.51, the estimated error covariance matrix was therefore the
Chapter 6. Application 2: Estimation of Tree Diameter Distribution

The following matrix.

\[
\hat{\Omega} = \begin{bmatrix}
5.2812W_1W'_1 & -1.1863W_1W'_2 & 0.4059W_1W'_3 \\
-1.1863W_2W'_1 & 0.8688W_2W'_2 & 0.2898W_2W'_3 \\
0.4059W_3W'_1 & 0.2898W_3W'_2 & 2.8225W_3W'_3
\end{bmatrix}
\] (6.172)

Because the \(a\) and \(b\) parameter equations have iid error terms, \(W_1\) and \(W_3\) are equal to the identity matrix of size \(n\) by \(n\). The \(W_2\) matrix is an \(n\) by \(n\) matrix with the diagonal elements equal to the square root of the estimated variances of the error terms for the second stage \(b\) parameter equation. The simplified error covariance matrix was therefore as follows:

\[
\hat{\Omega} = \begin{bmatrix}
5.2812I_n & -1.1863W_2 & 0.4059I_n \\
-1.1863W_2 & 0.8688W_2W_2 & 0.2898W_2W_2 \\
0.4059I_n & 0.2898W_2 & 2.8225I_n
\end{bmatrix}
\] (6.173)

6.6.5 EGLS to Fit the System of Equations

The last step of the MSLS fitting technique is to use the estimated error covariance matrix, and obtain an EGLS fit of the system of equations simultaneously. The resulting MSLS fit was as follows:

\[
pred. a_{MSLS} = 6.572920 + 0.076069 topht - 9.476689 \times 10^{-4} stems
+ 0.025394 totvol - 0.403223 pred. b_{1st}
\] (6.174)

\[
pred. b_{MSLS} = -6.095781 + 0.235510 qdiam + 0.001234 stems
- 0.028857 totvol + 1.235273 pred. a_{1st}
\] (6.175)

\[
pred. c_{MSLS} = 5.204059 - 0.299683 topht + 0.330314 pred. a_{1st}
+ 0.058220 pred. b_{1st}
\] (6.176)

The IMSLS technique was not used, because the number of samples was considered large enough for the asymptotic properties of the MSLS technique to be assumed.
6.7 Comparison of the Three Fitting Techniques

6.7.1 Goodness-of-fit Measures

Fit Index

The Fit Indices for each of the four equations for the OLS and the MSLS fits are presented in Table 6.15. The Fit Indices for the MSLS fit were lower than those of the OLS fit. The result was expected as the OLS fit minimizes the squared difference between the endogenous variable and the corresponding predicted value, whereas the MSLS fit minimizes the squared difference between weighted values. The differences between the OLS and the MSLS Fit Indices are larger for this second application than the differences shown for Application 1.

For the composite model, the Fit Index was calculated using the following equation.

$$FI = \frac{\sum_{m=1}^{n} (0.24 - F(D1))^2 + (0.63 - F(D2))^2 + (0.93 - F(D3))^2}{\sum_{m=1}^{n} (0.24 - 0.3333)^2 + (0.63 - 0.3333)^2 + (0.93 - 0.3333)^2}$$  (6.177)

where $F(D1)$, $F(D2)$, and $F(D3)$ are the estimated cumulative probabilities from the estimated Weibull distribution up to the $D1$, $D2$, and $D3$ diameter limits, respectively;

0.3333 is the mean of all three percentiles for the 121 plots;

0.24, 0.63, and 0.93 were the percentiles selected to represent each plot.

Table 6.15: Fit Indices for OLS and MSLS Fits of the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous</th>
<th>a parameter</th>
<th>b parameter</th>
<th>c parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.7383</td>
<td>0.6744</td>
<td>0.1563</td>
</tr>
<tr>
<td>MSLS</td>
<td>0.6398</td>
<td>0.5383</td>
<td>0.1057</td>
</tr>
</tbody>
</table>
In order to calculate $F(D_1)$, $F(D_2)$, and $F(D_3)$, the $a$ parameter was reset to 0.0 if a negative value was predicted. The Fit Index using this equation for the composite model fit was 0.6442. To compare to the OLS technique, the predicted values for the $a$, $b$, and $c$ parameter using the final OLS equations were used to obtain the estimated cumulative probabilities. One of the plots had an estimated $a$ parameter which was reset to 0.0, another had a negative $b$ parameter which was reset to $1.0 \times 10^{-7}$, and another had a negative $c$ parameter which was not reset. The overall Fit Index for the OLS technique, using the above equation, was 0.6679. Similarly, an overall Fit Index for the MSLS technique was calculated using predicted $a$, $b$, and $c$ parameters from the final MSLS fit. For one of the 121 plots, the predicted value of the $a$ parameter was negative, and was reset to 0.0. The overall Fit Index for the MSLS technique was 0.7125.

The overall Fit Index was lowest for the composite model fit which may have resulted because a local rather than a global minimum may have been found using NLP. Also, in the composite model, the parameters of the Weibull distribution are a function of the stand parameters alone and are not a function of the other parameters of the Weibull distribution. The MSLS overall Fit Index was slightly higher than the overall OLS Fit Index, which may be due the simultaneous fit of the three parameters using the MSLS technique. The presence of negative $b$ and $c$ parameters from the OLS fit may also have caused the lower overall Fit Index.

**Mean Absolute Deviation**

The mean absolute deviations (M.A.D.) were calculated by class for each of the LHS endogenous variables. The classes were created in a similar manner as for Application 1, with each class having 24 observations except the last class which had 25 observations. The M.A.D. values for the OLS fit and for the MSLS fit are presented in Table 6.16 and Table 6.17. As with the Application 1, the OLS technique resulted in lower M.A.D.
Chapter 6. Application 2: Estimation of Tree Diameter Distribution

Table 6.16: M.A.D. for Five Classes for the OLS Fit of the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>a parameter</td>
<td>1.884</td>
</tr>
<tr>
<td>b parameter</td>
<td>0.407</td>
</tr>
<tr>
<td>c parameter</td>
<td>1.174</td>
</tr>
</tbody>
</table>

Table 6.17: M.A.D. for Five Classes for the MSLS Fit of the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>a parameter</td>
<td>2.901</td>
</tr>
<tr>
<td>b parameter</td>
<td>0.557</td>
</tr>
<tr>
<td>c parameter</td>
<td>1.349</td>
</tr>
</tbody>
</table>
values across the range of the endogenous variables, although the differences between the OLS and MSLS M.A.D. values by class, were small.

For the composite model fit, an M.A.D. value was calculated for each percentile as shown below for the first percentile.

\[
M.A.D. = \frac{\sum_{m=1}^{n} |0.24 - F(D_1)|}{n}
\]  

(6.178)

Negative values for estimated parameters were reset as for the calculation of Fit Index. The M.A.D. for the 24th percentile was 0.190, for the 63rd percentile was 0.276, and for the 93rd percentile was 0.075. Using the estimated parameters from the final OLS fit to obtain \( F(D_1) \), \( F(D_2) \), and \( F(D_3) \), the M.A.D. for the 24th percentile was 0.236, for the 63rd percentile was 0.239, and for the 93rd percentile was 0.067. For the MSLS fit, the M.A.D. for the 24th percentile was 0.220, for the 63rd percentile was 0.223, and for the 93rd percentile was 0.070. The M.A.D. values were much the same for the three percentiles.

Mean Deviation

The mean deviations (M.D.) were calculated for the same classes as with the M.A.D. values. The results for the OLS fit are presented in Table 6.18, and for the MSLS fit in Table 6.19. As with the M.A.D. values by class, the M.D. values by class were lower with the OLS fit; however, the differences between the OLS and the MSLS fits are generally small. Also, the trends of over- and underestimation across the range of the endogenous variables are similar with the two fits.

For the composite model fit, an M.D. value was calculated for each percentile as shown for the first percentile.

\[
M.D. = \frac{\sum_{m=1}^{n} (0.24 - F(D_1))}{n}
\]

(6.179)
Table 6.18: M.D. for Five Classes for the OLS Fit of the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>a parameter</td>
<td>-0.727</td>
</tr>
<tr>
<td>b parameter</td>
<td>-0.293</td>
</tr>
<tr>
<td>c parameter</td>
<td>-1.079</td>
</tr>
</tbody>
</table>

Table 6.19: M.D. for Five Classes for the MSLS Fit of the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>a parameter</td>
<td>-1.421</td>
</tr>
<tr>
<td>b parameter</td>
<td>-0.529</td>
</tr>
<tr>
<td>c parameter</td>
<td>-1.349</td>
</tr>
</tbody>
</table>
Negative estimated parameters were reset as for the calculation of Fit Index. The M.D. for the 24th percentile was 0.175, for the 63rd percentile was 0.197, and for the 93rd percentile was -0.030. For the OLS fit, using the estimated Weibull parameters, the M.D. for the 24th percentile was 0.185, for the 63rd percentile was 0.033, and for the 93rd percentile was -0.054. The 24th and 63rd percentiles were therefore underestimated and the 93rd percentile was slightly overestimated. For the MSLS fit, the M.D. for the 24th percentile was 0.182, for the 63rd percentile was 0.035, and for the 93rd percentile was -0.064, much the same as for the OLS fit.

6.7.2 Relative Variances

The trace of the estimated coefficient covariance matrix for the MSLS fit was 49.64200, whereas the trace of the OLS fit was 4.7411 for the system of equations. Because the system of parameter equations was simultaneous, the OLS estimate of the coefficient covariance matrix cannot be used to calculate confidence limits for the true parameter. However, because the simultaneity bias was removed in the first step of the MSLS technique, the estimated coefficient covariance matrix from the MSLS fit is consistent. Confidence limits using these OLS variance estimates would appear, incorrectly, to be much narrower than the those using the consistently estimated coefficient covariance matrix from the MSLS fit.

The traces for the submatrices of the estimated coefficient covariance matrix corresponding to each individual equation are given in Table 6.20. Unlike Application 1, all of the equations of the dbh distribution system have endogenous variables on the RHS. The OLS estimates of the variance of the coefficients appear lower for all of the equations of the system, similar to the trace for the overall matrix.

The comparison of the coefficient covariance matrix for the composite model with those from the other fitting techniques was not possible, because the LHS variable of the
Table 6.20: Trace of the Coefficient Covariance Matrix for Each Equation of the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>a parameter</th>
<th>b parameter</th>
<th>c parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>3.690011</td>
<td>0.170987</td>
<td>0.880147</td>
</tr>
<tr>
<td>MSLS</td>
<td>5.056920</td>
<td>42.410241</td>
<td>1.174840</td>
</tr>
</tbody>
</table>

The composite model does not appear on the LHS of equations of the system fitted by the OLS or the MSLS procedure. Also, the variances were not calculated for the composite model.

6.7.3 Table of Estimated Coefficients and Standard Deviations

A summary of the estimated coefficients and their associated standard deviations from the OLS and MSLS fits is shown in Table 6.21. The estimated standard deviations for the coefficients were higher for the MSLS fit than for the OLS fit, but the OLS estimates of the coefficients are inconsistent, and so these standard deviations would result in underestimated confidence intervals for all of the coefficients in the system. The coefficients for the $a$ parameter and $c$ parameter equations were similar for the two techniques, but, for the $b$ parameter equation, the coefficients were quite different. Since the MSLS estimated coefficients are similar to the unweighted multiple least squares fit of each of the second stage equations (equations 6.164, 6.165, and 6.166), the difference in coefficients must be due to simultaneity bias.

Because the OLS results in biased coefficients, hypothesis statements tested using the results from the OLS fit will be incorrect. For instance, using a value of 1.96 from the normal distribution, confidence intervals for the $\delta_{15}$ coefficient which was associated with the $b$ parameter in the first equation, were calculated as -1.58437 to -0.86721 from
## Table 6.21: Estimated Coefficients and Standard Deviations of Coefficients for the Dbh Distribution System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Coefficients</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>MSLS</td>
</tr>
<tr>
<td><strong>a parameter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{11}$</td>
<td>7.21</td>
<td>6.57</td>
</tr>
<tr>
<td>$\delta_{12}$</td>
<td>0.26</td>
<td>0.08</td>
</tr>
<tr>
<td>$\delta_{13}$</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\delta_{14}$</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>$\delta_{15}$</td>
<td>-1.23</td>
<td>-0.40</td>
</tr>
<tr>
<td><strong>b parameter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{21}$</td>
<td>3.57</td>
<td>-6.09</td>
</tr>
<tr>
<td>$\delta_{22}$</td>
<td>0.16</td>
<td>0.24</td>
</tr>
<tr>
<td>$\delta_{23}$</td>
<td>-0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\delta_{24}$</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\delta_{25}$</td>
<td>-0.21</td>
<td>1.24</td>
</tr>
<tr>
<td><strong>c parameter</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{31}$</td>
<td>4.44</td>
<td>5.20</td>
</tr>
<tr>
<td>$\delta_{32}$</td>
<td>-0.34</td>
<td>-0.29</td>
</tr>
<tr>
<td>$\delta_{33}$</td>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td>$\delta_{34}$</td>
<td>0.54</td>
<td>0.06</td>
</tr>
</tbody>
</table>
the OLS fit, and as -1.26032 to 0.45388 from the MSLS fit. Using the results from the
MSLS fit, this coefficient would be considered to be zero, whereas for the OLS fit, the
coefficient is nonzero. Similar differences in hypothesis testing were noted for the $\delta_{25}$ and
$\delta_{34}$ coefficients.

### 6.7.4 Ranking for Other Features

The ranks assigned to each fitting technique are shown in Table 6.22. For the Infor-

<table>
<thead>
<tr>
<th>Feature</th>
<th>OLS</th>
<th>Composite</th>
<th>MSLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Consistency</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Confidence Limits</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Asymptotic Efficiency</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Compatibility</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ease of fit</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>9</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>

mation feature, the composite model was assigned a rank of one as only one endogenous
variable was estimated. The only consistent estimates were the MSLS coefficient esti-
mates, as the nonlinear least squares procedure used to fit the composite model was
restricted to obtain positive values for the predicted $b$ and for the $c$ parameter values.
Confidence limits can be calculated for MSLS as the sample size was quite large and the
estimated coefficients are asymptotically normally distributed. For the composite model,
confidence limits could be calculated if generalized nonlinear least squares were used to
fit the model, because estimates of the variances of the coefficients could be obtained. In
terms of efficiency, the MSLS approach should be more efficient than the OLS approach;
however, since the OLS estimates of all of the coefficients of this system were biased and inconsistent, this expected increase in efficiency using the MSLS approach was difficult to witness. The MSLS and composite model fits both result in compatibility, whereas the OLS fit does not. The number of steps for the MSLS fit is greater than that for the OLS fit; for the composite model, the number of steps was not calculated as no attempt was made to estimate the variances of the coefficients of the nonlinear composite model. Because the fit of the nonlinear composite model would require simultaneous estimation of the error covariance matrix and the coefficients if the error terms were non-iid, the composite model was considered to be the most difficult to fit. Also, the number of coefficients to be estimated in the nonlinear composite model was 15; the chance of obtaining a local minimum rather than a global minimum is probably high with this large number. The composite model was therefore given a low "ease of fit" ranking.

The MSLS technique was therefore assigned the highest sum of ranks for this second application.

6.8 Conclusion

The chosen system of parameter prediction equations was simultaneous. However, the LHS variable of the third equation, the c parameter, did not appear on the RHS.

One of the equations, the b parameter equation, was found to have heteroskedastic error terms using the OLS technique. The error terms for the c parameter equation indicated some lack-of-fit, but the error terms for this equation were assumed to be iid, as were the error terms for the a parameter equation. EGLS was used to fit the b parameter equation, whereas multiple linear regression was used to fit the other two equations.

For the composite model, a nonlinear model was derived with 15 coefficients to be
estimated. The equation was fitted using restricted nonlinear least squares. No attempt was made to check for serial correlation or heteroskedasticity.

The unweighted fit of the second stage equations, for the MSLS fit, indicated that the characteristics of the error terms were similar to those of the error terms from the unweighted fit for the OLS technique. The variances of the error terms for the second stage $b$ parameter equation were estimated and the equation was refitted using EGLS. The estimated error covariance matrix was then used to obtain the final MSLS fit.

The goodness-of-fit measures for the individual parameter prediction equations were best for the OLS fit, as expected. Unlike Application 1, the goodness-of-fit measures for the MSLS fit were somewhat worse than for the OLS fit. The goodness-of-fit measures for the composite model were compared to the simulated composite model using the OLS fit, and, separately, the MSLS fit. The goodness-of-fit measures, in this case, were generally best for the MSLS fit, except for the M.D. values; however, since the OLS fit resulted in one negative predicted $b$ parameter and one negative predicted $c$ parameter, the goodness-of-fit measures may be somewhat misleading for the simulated composite model. The goodness-of-fit measures for the composite model were the worst.

The trace of the estimated coefficient covariance matrix was lowest with the OLS fit. The trace values for the submatrices of this matrix, corresponding to each equation of the system, were lower for OLS than for MSLS. Since all three of the parameter prediction equations have RHS endogenous variables, these results are similar to those found for the total volume and volume ratio equation of Application 1 which also had RHS endogenous variables. Since the OLS fit coefficients are biased and inconsistent, these estimated variances can not be used to obtain confidence limits for the true coefficients.

The summed rank for other features was highest for the MSLS fit. The composite model was given a low ease-of-fit rank because the simultaneous estimation of the 15 coefficients was difficult, and if estimates of the variances of these coefficients is also
desired, the estimation would be even more difficult. The OLS fit resulted in biased and inconsistent estimates as all of the equations have RHS endogenous variables.

In terms of goodness-of-fit, the OLS fit was better; if a fit of the sample data were required and this fit was not to be used for other samples of the population, the OLS fit would be appropriate. However, the OLS fit results in inconsistent estimates of the coefficients, and the results could not be used to test hypothesis statements and should not be used for other sample data. The composite model fit would require a difficult fitting procedure to obtain estimates of the coefficient covariance matrix, and the goodness-of-fit measures were lower than those of the MSLS fit for the simulated composite model. Because tests of hypothesis statements are useful, and the fitted equation may be used for other samples of the population, the MSLS approach was considered the best technique for fitting this system of equations; the coefficients and the estimated variances of the coefficients are consistent.
Chapter 7

Application 3: Volume Growth and Yield

7.1 Introduction

The management of forest resources requires accurate information about the current and future wood supply. Systems of equations have been widely used to represent forest growth and yield. Simultaneous fitting techniques such as 2SLS, 3SLS, restricted 3SLS, and minimizing loss functions for the system have been used to fit these equations (Borders and Bailey, 1986; Burkhart and Sprinz, 1984; Furnival and Wilson, Jr., 1971; Hans, 1986; Murphy and Beltz, 1981; Murphy and Sternitzke, 1979; Reed et al., 1986). However, for each of these studies, the assumption was made that individual equations of the system have iid error terms.

Permanent sample plots are often used in fitting growth and yield systems. Because psps are measured repeatedly over time, error terms may be serially correlated, and also, the variance of the error terms may differ among plots. The presence of serial correlation between error terms will depend on the length of time between measurements and whether overlapping intervals are used in estimation (Borders et al., 1987a). The use of simultaneous fitting techniques for systems of growth and yield equations assuming that the error terms are iid may therefore be less efficient than the MSLS technique, depending on the degree of serial correlation or heteroskedasticity of the error terms.
7.2 Data Preparation

The psp data used for Application 2 were also selected for this application. Plots with more than 80 percent pine by volume, breast height age recorded, and no treatment applied were selected. Data from Northeastern Alberta were deleted so that only lodgepole pine was represented in the data. One plot was selected from each cluster of four plots at a location. Summary information from the establishment measurement and the subsequent measurement was selected; any plot having only the establishment measurement was deleted from the data. Data were graphed and no outliers were noted. The distribution of the remaining 28 plots is given in Table 7.23.

Table 7.23: Distribution of Selected Psp's for the Third Application

<table>
<thead>
<tr>
<th>Stems per Hectare</th>
<th>Age at Breast Height</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21 to 40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41 to 60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>61 to 80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>81 to 100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>101 to 120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>121 to 140</td>
<td></td>
</tr>
<tr>
<td>0 to 1000</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1001 to 2000</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2001 to 3000</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3001 to 4000</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4001 to 5000</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5001 to 6000</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6001 to 7000</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

|                  | 5                    | 2     |
|                  | 6                    | 3     |
|                  |                      | 28    |
Chapter 7. Application 3: Volume Growth and Yield

7.3 Model Selection

Two equations of the growth and yield system developed by Clutter (1963) were selected for analysis. These equations were as follows:

\[ \ln BA_2 = \ln BA_1 \frac{A_1}{A_2} + \alpha_1 \left( 1 - \frac{A_1}{A_2} \right) + \alpha_2 \left( 1 - \frac{A_1}{A_2} \right) SI + \text{error}_1 \]  
\[ \ln V_2 = \beta_0 + \beta_1 SI + \beta_2 \frac{1}{A_2} + \beta_3 \ln BA_2 + \text{error}_2 \]

where \( \ln BA_1 \) and \( \ln BA_2 \) are the natural logarithms of the basal area per hectare measured at times 1 and 2, respectively;

\( A_1 \) and \( A_2 \) are the number of years counted at breast height at times 1 and 2, respectively;

\( SI \) is the site index for a reference age of 50 years measured at breast height;

\( \ln V_2 \) is the natural logarithm of volume per hectare at time 2;

\( \alpha_1 \) and \( \alpha_2 \), and \( \beta_0 \) through \( \beta_3 \) are coefficients to be estimated;

\( \text{error}_1 \) and \( \text{error}_2 \) are the error terms.

These equations were fitted simultaneously by Burkhart and Sprinz (1984) using a minimum loss function for the two equations. Borders and Bailey (1986), Murphy and Ster- nitzke (1979), and Hans (1986) used simultaneous fitting techniques from econometrics, assuming that error terms of individual equations are iid, to fit modifications of these equations.

The equations can be rearranged as follows:

\[ \left( \ln BA_2 - \ln BA_1 \frac{A_1}{A_2} \right) = \delta_{11} + \delta_{12} \left( 1 - \frac{A_1}{A_2} \right) + \delta_{13} \left( 1 - \frac{A_1}{A_2} \right) SI + \epsilon_1 \]  
\[ \ln V_2 = \delta_{21} + \delta_{22} SI + \delta_{23} \frac{1}{A_2} + \delta_{24} \ln BA_2 + \epsilon_2 \]

The term \( \ln BA_1 \frac{A_1}{A_2} \) is a combined term where \( BA_1 \) is a lagged endogenous variable which can be treated as a predetermined variable, and the term \( \frac{A_1}{A_2} \) is an exogenous
variable. This combined term can be treated as a constant which is subtracted from the
endogenous variable, \( \ln BA_2 \), to obtain the LHS endogenous variable shown in the basal
area equation above. An intercept coefficient was added to this basal area equation,
although this coefficient is expected to be close to zero.

7.4 Ordinary Least Squares Fit

7.4.1 Unweighted Simple or Multiple Linear Regression

The estimated coefficients using multiple linear regression to fit each equation of the
system individually were as follows:

\[
\text{pred. } (\ln BA_2 - \ln BA_1 \frac{A_1}{A_2}) = -0.017785 + 4.458907 \left(1 - \frac{A_1}{A_2}\right) - 0.011270 \left(1 - \frac{A_1}{A_2}\right) SI \tag{7.184}
\]

\[
\text{pred. } \ln V_2 = 2.315771 + 0.066766 SI - 47.758153 \frac{1}{A_2} + 1.014053 \ln BA_2 \tag{7.185}
\]

The estimated error terms from the unweighted fit were then used to test for serial
correlation and heteroskedasticity.

7.4.2 Testing for iid Error Terms

To check for serial correlation in each equation, the estimated error terms were sorted
by the predicted endogenous variables, and a graph of the current error term versus the
previous error term was obtained. The \( R^2 \) value for the simple linear regression of the
current error term with the previous error term was 0.00965 for the basal area equation
and 0.011511 for the volume equation. The Durbin and Watson (1951) test statistic for
each model was not significant for either positive or negative serial correlation using an
alpha of 0.10 (0.05 for positive and 0.05 for negative serial correlation). Each equation of the system therefore had independent error terms.

To test for heteroskedasticity, a graph of the estimated error versus the predicted LHS endogenous variable was done. The regression of the estimated error, squared, with the predicted variable on the LHS of the basal area equation resulted in an \( R^2 \) value of 0.00270. The error terms for this equation were therefore considered identically distributed. The regression of the estimated error, squared, with the predicted logarithm of volume from the second equation, resulted in an \( R^2 \) value of 0.01425. The test statistic for the Goldfeld and Quandt (1965) test was 1.9149 for nonincreasing variance with increasing predicted logarithm of volume. The critical F value for 6 degrees of freedom for the numerator and the denominator, and an alpha level of 0.05 was 4.28. The error terms of the second equation of the system were therefore considered identically distributed.

### 7.4.3 Appropriate OLS Fit Based on Error Structure

Because each of the equations of the system had iid error terms, the unweighted OLS fit was considered appropriate. This unweighted fit was therefore used in the comparison with other methods.

### 7.5 Composite Model Fit

#### 7.5.1 Derivation of the Composite Model

A composite model was derived by Sullivan and Clutter (1972), by substituting the equation for the logarithm of basal area at time 2 into the equation to estimate yield at time two. The resulting equation was as follows:

\[
\ln V_2 = \gamma_0 + \gamma_1 SI + \gamma_2 \frac{1}{A_2} + \gamma_3 \ln BA_i \frac{A_3}{A_2} + \gamma_4 \left(1 - \frac{A_i}{A_2}\right)
\]
where $\gamma_0$ through $\gamma_5$ are coefficients to be estimated;

$\epsilon_3$ is the error term.

### 7.5.2 Unweighted Regression of the Composite Model

All of the RHS variables of the composite model can be considered predetermined variables and are uncorrelated with the error term. The fit using multiple linear regression results in unbiased estimates of the coefficients and their variances. The fitted equation was as follows:

$$
\text{pred. ln } V_2 = 2.513266 + 0.063926 SI - 39.294313 \left(1 - \frac{A_1}{A_2}\right) - 0.961204 \ln BA_1 \frac{A_1}{A_2} + 3.655157 \left(1 - \frac{A_1}{A_2}\right) - 0.003875 \left(1 - \frac{A_1}{A_2}\right) SI
$$

### 7.5.3 Testing for iid Error Terms

Data were ordered by the predicted logarithm of volume using the unweighted fit of the composite model, and a graph of the estimated error with the previous estimated error term was examined for serial correlation. The regression of the current error term with the previous error term resulted in an $R^2$ value of 0.0082, and the Durbin and Watson test statistics for positive and for negative serial correlation were not significant for an alpha of 0.10 for both tests. The error terms for the composite model were therefore independent.

To test for heteroskedasticity, a graph of the estimated error term with the predicted logarithm of volume was obtained, and the simple linear regression of the estimated error term, squared, with the predicted logarithm of volume was performed. The $R^2$ value was 0.03734. The Goldfeld and Quandt (1965) test statistic was 1.0567 assuming
that the variance is nonincreasing with increasing predicted logarithm of volume. This
test statistic was less than the critical F value of 5.05 for 5 degrees of freedom for the
numerator and for the denominator, and for an alpha of 0.05. The error terms were
therefore considered homoskedastic.

7.5.4 Weighted Fit of the Composite Model

The error terms for the composite model were found to be iid. The unweighted fit was
therefore used to compare to other fitting techniques.

7.6 MSLS Fit

7.6.1 First Stage Equations

The two equation of the system met the rank and order conditions and were identified.
For the second equation of the system, \( \ln BA_2 \) appears on the RHS and is the part of
the complex variable which appears as the endogenous variable on the LHS of the first
equation. To remove the simultaneity bias for the second equation, the following first
stage model was obtained.

\[
pred. \left( \ln BA_2 - \ln BA_1 \frac{A_1}{A_2} \right)_{1st} = 0.163469 - 0.013701 SI - 1.026070 \frac{1}{A_2} + 4.071904 \\
\quad \left( 1 - \frac{A_1}{A_2} \right) + 0.026270 \left( 1 - \frac{A_1}{A_2} \right) SI \tag{7.188}
\]

The predicted \( \ln BA_2 \) values were then recovered from this first stage equation using the
following equation.

\[
pred. \ln BA_{21st} = pred. \left( \ln BA_2 - \ln BA_1 \frac{A_1}{A_2} \right)_{1st} + \ln BA_1 \frac{A_1}{A_2} \tag{7.189}
\]
7.6.2 Second Stage Equations

The second stage equations were derived by substituting the predicted value for the logarithm of basal area at time 2 from the first stage equation into the RHS of the \( \ln V_2 \) equation. Because the second stage equation for \( \ln BA_2 - \ln BA_1 \frac{A_1}{A_2} \) has only exogenous variables on the RHS, the unweighted OLS fit of this equation was the same as the fit of the second stage equation.

\[
pred. \left( \ln BA_2 - \ln BA_1 \frac{A_1}{A_2} \right)_{2nd} = -0.017785 + 4.458907 \left( 1 - \frac{A_1}{A_2} \right) - 0.011270 \left( 1 - \frac{A_1}{A_2} \right) SI \tag{7.190}
\]

\[
pred. \ln V_{2nd} = 2.550358 + 0.066408 SI - 47.37629 \frac{1}{A_2} + 0.9503109 pred. \ln BA_{21st} \tag{7.191}
\]

7.6.3 Testing for iid Error Terms

The basal area second stage equation was the same as the basal area equation for the unweighted OLS fit; therefore, the error terms for this second stage equation were iid.

To test for serial correlation in the second stage volume model, the estimated error terms were ordered by the predicted logarithm of volume, and a graph of the estimated error term with the previous error term was obtained. The \( R^2 \) value for the simple linear regression of the current with the previous error term was 0.01856. The Durbin and Watson (1951) test statistics for positive and negative serial correlation were not significant for an alpha of 0.10 for the two sided test. The error terms were therefore considered independent for the second stage logarithm of volume equation.

To test for heteroskedasticity, a graph of the estimated error with the predicted logarithm of volume using the second stage model was obtained. The regression of the estimated error, squared, with predicted logarithm of volume resulted in an \( R^2 \) value
of 0.0950. The Goldfeld and Quandt (1965) test for an alpha of 0.05 was not significant. The error terms of the second stage logarithm of volume equation were therefore identically distributed.

7.6.4 Estimation of the Error Covariance Matrix

Because the equations of the system have iid error terms, the MSLS procedure simplifies to the 3SLS procedure. Estimates of contemporaneous variances using equation 3.28 were calculated from the estimated error terms of the second stage models, resulting in the following $\hat{\Sigma}$ matrix.

$$
\hat{\Sigma} = \begin{bmatrix}
0.0067 & 0.0060 \\
0.0060 & 0.0078
\end{bmatrix}
$$

(7.192)

The error covariance matrix for the system was therefore as follows:

$$
\hat{\Omega} = \begin{bmatrix}
0.0067I_n & 0.0060I_n \\
0.0060I_n & 0.0078I_n
\end{bmatrix}
$$

(7.193)

The covariance of the error terms between equations 0.0060 was quite high relative to the variance of the error terms within each equation which were 0.0067 and 0.0078. The equations therefore appear to be correlated.

7.6.5 EGLS to Fit the System of Equations

The final step of the MSLS fit was to fit the system simultaneously using the estimated error covariance matrix, $\hat{\Omega}$. The resulting MSLS fit of the system of equations was as follows:

$$
pred. \left( \ln BA_2 - \ln BA_1 \frac{A_1}{A_2} \right)_{MSLS} = -0.0480136 + 4.588017 \left(1 - \frac{A_1}{A_2}\right)
-0.008839 \left(1 - \frac{A_1}{A_2}\right) SI
$$

(7.194)
Chapter 7. Application 3: Volume Growth and Yield

\[
\text{pred.} \ln V_{2MSLS} = 2.264741 + 0.072275 SI - 47.47299 \frac{1}{A_2} + 1.004244 \text{pred.} \ln BA_{21st} \tag{7.195}
\]

The IMSLS technique in this case is the same as the I3SLS technique. The estimated coefficients from the MSLS fit were used to estimate new values for the contemporaneous correlation by using equation 3.32, and a new MSLS fit was calculated. This process was repeated for eight iterations, until the criterion specified in equation 4.65 was met. Coefficients were similar to those for the MSLS fit.

7.7 Comparison of the Three Fitting Techniques

7.7.1 Goodness-of-fit Measures

Fit Index

The Fit Indices for each of the two equations for the OLS, MSLS, and IMSLS fits are presented in Table 7.24. Fit Indices for the composite model fit are presented for the logarithm of volume only. The Fit Indices for the MSLS fit were marginally lower than those for the OLS fit, for the basal area equation. For the volume equation, the Fit Index for the OLS fit was the highest, followed by the composite model, and the lowest Fit Index was for the IMSLS fit.

Table 7.24: Fit Indices for OLS, Composite Model, MSLS and IMSLS Fits of the Yield Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>OLS</th>
<th>Composite Model</th>
<th>MSLS</th>
<th>IMSLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln BA_2 - ln BA_1 A_1</td>
<td>0.9536</td>
<td>none</td>
<td>0.9519</td>
<td>0.9496</td>
</tr>
<tr>
<td>ln V_2</td>
<td>0.9802</td>
<td>0.9129</td>
<td>0.9024</td>
<td>0.8974</td>
</tr>
</tbody>
</table>
Mean Absolute Deviation

The mean absolute deviations (M.A.D.) were calculated by class for each of the LHS endogenous variables. The classes were created by sorting the 28 samples by the endogenous variable, and then dividing the sorted data into four classes of six samples each, with the fifth class having only four samples. The M.A.D. values for the OLS fit, for the MSLS fit, and for the composite model fit for the logarithm of volume only, are presented in Table 7.25, Table 7.26, and Table 7.27. No values are shown for the IMSLS fit; the values were within 0.005 units from the MSLS fit.

Table 7.25: M.A.D. for Five Classes for the OLS Fit of the Yield Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\ln BA_2 - \ln BA_1 \frac{A_1}{A_2}$</td>
<td>0.043</td>
</tr>
<tr>
<td>$\ln V_2$</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Table 7.26: M.A.D. for Five Classes for the MSLS Fit of the Yield Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\ln BA_2 - \ln BA_1 \frac{A_1}{A_2}$</td>
<td>0.039</td>
</tr>
<tr>
<td>$\ln V_2$</td>
<td>0.090</td>
</tr>
</tbody>
</table>

For the basal area equation, the M.A.D. values were only marginally lower for the OLS fit compared to the MSLS fit. For the volume equation, the M.A.D. values were somewhat higher for the MSLS and composite model fits, relative to the OLS fit. The
Table 7.27: M.A.D. for Five Classes for the Yield Composite Model Fit

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln V_2 )</td>
<td>1      2   3    4   5</td>
</tr>
<tr>
<td>0.089</td>
<td>0.073  0.039 0.080 0.057</td>
</tr>
</tbody>
</table>

The composite model fit had similar M.A.D. values to the MSLS fit.

**Mean Deviation**

The mean deviations (M.D.) were calculated for each equation for the same five classes as for M.A.D. Results for the OLS fit are presented in Table 7.28, for the MSLS fit in Table 7.29, and for the Composite Model fit in Table 7.30. The M.D. values for the IMSLS fit were similar to the MSLS fit for the volume equation. For the basal area equation, values for the IMSLS fit were close those for the OLS fit.

Table 7.28: M.D. for Five Classes for the OLS Fit of the Yield Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln BA_2 - \ln BA_1 )</td>
<td>1      2   3    4   5</td>
</tr>
<tr>
<td>-0.024</td>
<td>-0.010 0.008 -0.004 0.044</td>
</tr>
<tr>
<td>( \ln V_2 )</td>
<td>-0.003 -0.017 0.003 0.002 0.023</td>
</tr>
</tbody>
</table>

For the basal area equation, the M.D. values were similar for the OLS and MSLS fits. The directions of the deviations between the observed and predicted values were similar also. For the volume equation, the M.D. values were lower for the OLS fit than for the MSLS fit. The M.D. values for the composite model fit were the highest.
Table 7.29: M.D. for Five Classes for the MSLS Fit of the Yield Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln BA₂ - ln BA₁ A₁ / A₂</td>
<td>1</td>
</tr>
<tr>
<td>ln V₂</td>
<td>-0.006</td>
</tr>
</tbody>
</table>

Table 7.30: M.D. for Five Classes for the Yield Composite Model Fit

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>Classes, from Low to High Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln V₂</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

7.7.2 Relative Variances

The trace of the estimated coefficient covariance matrix for the MSLS fit was 12.35046 whereas the trace of the OLS fit was 7.500340 for the system of equations. However, the OLS fit results in coefficients which are inconsistent, because the endogenous variable on the LHS of the first equation was modified and included as an endogenous variable on the RHS of the second equation of the system.

The traces for the submatrices of the estimated coefficient covariance matrix corresponding to each individual equation using the OLS and MSLS fitting techniques are given in Table 7.31. The trace of the coefficient covariance matrix for the composite model fit is shown in the table; however, the composite model had six coefficients in the volume equation, whereas for the OLS and MSLS fit of the system of equations, only four coefficients were estimated for the volume equation.
Table 7.31: Trace of the Coefficient Covariance Matrix for Each Equation of the Yield Equation System

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>OLS</th>
<th>MSLS</th>
<th>Composite Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln BA_2 - \ln BA_1 \frac{A_1}{A_2}$</td>
<td>0.59786</td>
<td>0.21083</td>
<td>none</td>
</tr>
<tr>
<td>$\ln V_2$</td>
<td>6.9025</td>
<td>12.1396</td>
<td>98.1090</td>
</tr>
</tbody>
</table>

The high trace value for the composite model was attributable mostly to the high variance of the coefficient associated with the $\frac{1}{A_2}$ variable. Also, the mean squared error for the composite model was 0.00933, whereas for the OLS fit of the volume equation, the mean squared error was only 0.00194. The higher mean squared error and larger number of variables likely caused the high value for the trace of the estimated coefficient covariance matrix from the composite model.

The OLS fit of the basal area equation resulted in unbiased estimates of the coefficients and their variances, because all RHS variables of this equation were exogenous. The MSLS fit resulted in a lower trace value for the estimated coefficient covariance matrix, which was the result of using the information from the volume equation in fitting the basal area equation (contemporaneous variances). The OLS fit of the volume equation results in biased and inconsistent coefficients, because basal area appears as a LHS and as a RHS variable in the system of equations. The trace of the estimated coefficient covariance matrix was therefore lower than for the MSLS fit. Because the number of samples used to fit this system of equations was small, the large sample properties of consistency and asymptotic efficiency may not apply for the 3SLS fit. The OLS fit has been shown to be more biased than 3SLS, and since the MSLS fit for this system was simply a 3SLS fit, the OLS results may be more biased than those of the MSLS fit, for the volume equation.
7.7.3 Table of Estimated Coefficients and Standard Deviations

A summary of the estimated coefficients and their associated standard deviations from the OLS and MSLS fits is shown in Table 7.32.

Table 7.32: Estimated Coefficients and Standard Deviations of Coefficients for the Yield Equation System

| Table 7.32: Estimated Coefficients and Standard Deviations of Coefficients for the Yield Equation System |
|-------------------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Endogenous Variable                             | Coefficients                   | Standard Deviation              |                                      |
|                                                | OLS                             | MSLS                            | IMSLS                             |
| \(\ln B A_2 - \ln B A_1 A_3\)                 | -0.017785                      | -0.048014                       | -0.059548                         | 0.048789                        | 0.034473                        | 0.032591                        |
| \(\delta_{11}\)                                | 4.458907                       | 4.588017                        | 4.561855                          | 0.770864                        | 0.453306                        | 0.394548                        |
| \(\delta_{12}\)                                | -0.011270                      | -0.008839                       | -0.003256                         | 0.035310                        | 0.022136                        | 0.019709                        |
| \(\ln V\)                                      | 2.315771                       | 2.264741                        | 2.202168                          | 0.134221                        | 0.163504                        | 0.145804                        |
| \(\delta_{21}\)                                | 0.006766                       | 0.072275                        | 0.074283                          | 0.005269                        | 0.006985                        | 0.006259                        |
| \(\delta_{22}\)                                | -47.758153                     | -47.47299                       | -46.69858                         | 2.623611                        | 3.480158                        | 3.119119                        |
| \(\delta_{23}\)                                | 1.104053                       | 1.004244                        | 1.010136                          | 0.033243                        | 0.036704                        | 0.031243                        |

The estimated coefficients were similar for the three techniques. Because the IMSLS technique was simply I3SLS for this problem, and the I3SLS should converge to MLE if the error terms are normally distributed, the I3SLS estimates of the coefficients may be more appealing.

7.7.4 Ranking for the Three Techniques

The ranks assigned to each fitting technique are shown in Table 7.33. Because of the small number of samples, the assumption that the MSLS estimates are normally distributed may not be applicable; the calculation of confidence limits using the results from this small number of samples would likely be incorrect. The MSLS estimators are
consistent, but, because of the small number of samples, the estimates obtained may be quite different from the true values.

The summed rank for these features was highest for the composite model, because estimates of the coefficients are unbiased and so are the estimated variances of these coefficients. However, the composite model only estimates the coefficients for the endogenous variable of the second equation of the system. For the first equation, the basal area equation, the coefficients from the fitted volume equation may be used, as demonstrated by Sullivan and Clutter (1972), but these estimated coefficients are not necessarily unbiased or consistent estimates.

The OLS fit resulted in inconsistent estimates of the coefficients for the volume model, as one of the RHS variables is endogenous. Also, the ease-of-fit rank was lower than for the composite model fit, as only one equation is fitted for the composite model. Finally, the OLS fit does not result in compatible equations.
7.8 Conclusion

The system of equations proposed by Clutter (1963) was selected for this application. This system is simultaneous, although the first equation has only exogenous variables on the RHS.

The estimated error terms from the unweighted OLS fit were iid, even though two measurements from psp's were used to analyze the two equation yield system, and serial correlation was expected to be significant. This may have resulted because the plot data were pooled and serial correlation within measurements of a plot was masked by this grouping of data. Also, the measurement period between the first and second varied from five to 14 years, and the longer periods result in a reduction in the correlation between measurements. The unweighted fit was used for comparison with the other fitting techniques.

The composite model was derived by Sullivan and Clutter (1972). The estimated error terms from the multiple linear regression of the composite model were iid.

The estimated error terms from the unweighted fit of the second stage equations for the MSLS fit, were iid. The MSLS technique therefore was reduced to the 3SLS technique. The estimated contemporaneous variance between the two equations was high relative to the estimated variances with each equation and the error terms between equations appeared to be highly correlated.

The goodness-of-fit measures were generally best for the OLS fit. For the basal area equation, the MSLS goodness-of-fit measures were close to those of the OLS fit, and sometimes better. For the volume equations, the goodness-of-fit measures for the composite model fit and for the MSLS fit were similar, and somewhat worse than those of the OLS fit.

The trace of the estimated coefficient covariance matrix was lower for the OLS fit than
for the MSLS fit. An examination of the trace values for the submatrices indicated that, like Application 1, the trace value for the equation with no RHS endogenous variables (basal area equation) was lower for the MSLS fit than for the OLS indicating an increase in efficiency by accounting for contemporaneous correlation. Like Applications 1 and 2, the trace value for the submatrix of the estimated coefficient covariance matrix for the equation with RHS endogenous variables, the volume equation, was lower for the OLS fit. If the variances of the coefficients from the OLS fit were used to calculate confidence limits, the confidence interval would be underestimated.

The composite model fit was given the highest summed rank for other features. This was largely due to the small number of samples used to test this application. The MSLS technique was given a lower rank because some of the properties of EGLS could not be applied to the MSLS fit for only 28 samples. If the number of samples was larger, the summed rank for the MSLS fit would be equal to that of the composite model fit.

The composite model only estimated the LHS variable; however, unbiased estimates of the coefficients were obtained. In terms of goodness-of-fit measures, the OLS fit was best, but the estimates of the coefficients of the volume model were biased and inconsistent. Also, information concerning the contemporaneous correlation was not used for the OLS fit, and so the estimates are less efficient than with the MSLS technique. The MSLS fit was not necessarily the best for this application, because of the small number of samples used. The large sample properties of EGLS, applicable to MSLS could not be assumed. If more two-measurement psps were available, the large sample properties of the MSLS technique could be assumed and so this technique would be the most favorable. Alternatively, all possible pairs of the measurements from the selected psps could have been used to fit the model, but serial correlation between the measurements of each plot may have been significant. Using all possible pairs of measurements, the error structure would likely be more complex than that given for the serially and contemporaneously
correlated systems of equations in Chapter 3 of this thesis.

If compatibility and estimates of each of the LHS variables of this system were re­
quired, the IMSLS procedure would be the most appropriate, because these estimates
would be equal to the MLE estimates if error terms were normally distributed. The
standard deviations shown for the IMSLS fit were the lowest for the basal area equation,
and were lower than those of the MSLS fit for the volume equation.
Chapter 8

Overall Discussion and Conclusions

The first hypothesis of this thesis was that a fitting technique exists which satisfies the desired criteria for simultaneous, contemporaneously correlated systems of equations, in which individual equations have non-iid error terms. The second hypothesis was that any additional computational burden in using the technique is compensated by the benefits of meeting the desired criteria.

The first objective of this research, related to the first hypothesis, was to find a technique from the literature which meets the desired criteria for simultaneous, contemporaneously correlated systems of equations, in which the error terms for individual equations are non-iid. This objective was not met, because no technique was found which satisfied these criteria for systems of equations with this error structure. However, information from the literature was used to derive a new fitting technique, labelled multistage least squares (MSLS), which is an extension of the 3SLS technique to systems in which the error terms of individual equations are non-iid. The estimated coefficients from the MSLS technique are consistent and asymptotically efficient, if the estimated error covariance matrix is consistent and the error structure has been correctly determined for the system. Confidence limits can be calculated for large sample sizes, and compatibility is maintained.

The second objective, related to the second hypothesis, was to compare the chosen technique to the common techniques of OLS applied to each equation, and OLS applied
to a composite model. Since no technique was found in literature, the MSLS technique was used for this comparison. The three techniques were applied to three forestry problems for which systems of equations are used. The criteria for examining the results of the three techniques included the comparison of goodness-of-fit measures (Fit Index, Mean Absolute Deviation, Mean Deviation), the comparison of the trace of the estimated coefficient covariance matrix, and the calculation of a summed rank based on the amount of information given, the consistency of estimates using information from literature to assess the system, the ability to calculate confidence intervals, the efficiency using information from literature to assess the system, the compatibility, and the ease of fit in terms of the number and difficulty of steps required.

The OLS fit of individual equations is simple to calculate and algorithms are readily available. The OLS fit of each of the systems of equations for the three applications, resulted in better goodness-of-fit measures than did the MSLS fit, as expected, because the OLS fit minimizes the sum of squared differences. Also, the estimated coefficients from the OLS fit were generally close to those from the MSLS fit. The OLS fit requires less computer memory than the MSLS fit; large forestry problems with many equations, variables, and samples can be fitted. However, for simultaneous systems of equations, the estimated coefficients are biased and inconsistent. The estimates do not converge to the true estimates, with increasing sample size. Also, confidence limits cannot be calculated and compatibility within the system is not assured.

The OLS fit of a composite model, created by substituting all of the equations into one equation, was simple to perform for the applications tested. Also, for the two linear composite models, the estimated coefficients were unbiased, because all of the RHS variables were exogenous. For the nonlinear composite model, coefficients were restricted and are therefore likely biased. For the first application, the volume system, the derived
Chapter 8. Overall Discussion and Conclusions

composite model did not appear to have all of the important variables. For the dbh distribu­tion, the composite model did not show the relationships among the parameters of the Weibull distribution. Because the composite model derived for the growth and yield system is useful for predicting only the volume yield, the basal area yield was not predicted. The technique of deriving and fitting a composite model meets all of the desired criteria for some systems of forestry equations. However, if endogenous variables remain on the RHS, estimated coefficients are biased and consistent. Since only one endogenous variable is predicted, the composite model fit may be undesirable. Also, the original biological relationships may be lost, and important variables may not be retained in the derived model. Finally, for large problems with many variables and many equations, this technique is impractical.

For all three of the applications tested, the goodness-of-fit measures for the MSLS fit were close to those for the OLS fit, and were sometimes better than the composite model fit. The number of steps required was similar to the OLS fit, also. In addition, the estimated coefficients of the MSLS fit were consistent and asymptotically efficient, except for the first application for which some efficiency was lost. Compatibility was also obtained with the MSLS fit. Hypothesis statements can also be tested. For the first and second applications, the use of the OLS fit to incorrectly test hypotheses about coefficients resulted in different conclusions than if the consistent estimates from the MSLS fit were used. The selected applications did not demonstrate the use of the MSLS technique for serially correlated error terms in individual equations, or for heteroskedastic, serially correlated error terms. However, the desired criteria would still be met for these error structures, and the difficulty in obtaining the MSLS fit would likely be similar to the applications tested in this thesis.

The main disadvantage of the MSLS fit was that more computer memory is required,
than for the OLS or composite model fits. For large forestry problems with many equations, variables, and samples, a more efficient computer program or enough computer memory would be required. Alternatively, a modified MSLS technique could be used, in which the final step of the MSLS technique is not performed. This modified MSLS technique is simply an extension of the 2SLS technique to non-iid error terms. Estimates of coefficients and their variances would remain consistent for this modified MSLS technique. A loss in efficiency would be incurred, because the information about contemporaneous correlation used in the last step of the MSLS technique would not be utilized. The final alternative, which was used for Application 1, is to fit the system using a subsample of the data. Again, a loss of efficiency would result.

Another disadvantage of the MSLS technique is that estimates are not unbiased. For small samples, a Monte Carlo study to examine the degree of bias for small sample sizes for the MSLS and IMSLS techniques should be conducted. The contemporaneous correlation, correlation of the RHS variables with the error term, serial correlation, and heteroskedasticity should be varied to examine the effects on bias. However, studies using 3SLS for small samples indicated that the bias is often less for this technique than for the OLS technique. Similar results may occur for the MSLS technique.

In summary, the first hypothesis of this thesis was refuted, because no technique was found in literature which meets the desired criteria for the error structure described. However, a fitting technique was derived as part of this research. In terms of the number of steps required, the second hypothesis was met using MSLS for the applications tested. But in terms of the computer memory required, the MSLS technique results in an additional computational burden which could limit the use. A modified MSLS technique or a subsample of the data followed by the MSLS technique are proposed as alternatives.

A FORTRAN program with International Math and Statistical Library (IMSL) subroutines, Version 1.0, was used to obtain the final MSLS fits for the three applications presented in this thesis.
for large forestry problems. Also, since the computer program used in this research may not be most efficient, the computational burden may be reduced by creating an efficient routine.
Chapter 9

References Cited


160


Pap. NC-36. 8 p.


Chapter 9. References Cited


Appendix A

Glossary of Terms and Abbreviations

2SLS Two Stage Least Squares; a technique to fit systems of equations which are simultaneous but not contemporaneously correlated.

3SLS Three Stage Least Squares; a technique to fit systems of equations which are simultaneous and contemporaneously correlated.

AFS Alberta Forest Service

BLUE Best Linear Unbiased Estimator

compatibility Logical relationships between equations of a system are retained in the fitted system of equations.

composite All equations of the system are combined into one equation.

consistent The probability that the estimate is within a small deviation from the population value approaches one as the number of samples is increased; the estimate converges to the true value as the sample size is increased.

contemporaneous correlation Correlation of error terms between equations of a system.

contemporaneous variances The variances corresponding to contemporaneous correlation.

dbh diameter at breast height (1.3 metres above ground)

efficient The Cramer-Rao lower bound for efficiency is met. More efficient means that the variance of the estimate is lower whereas less efficient means that the variance of the estimate is higher.

EGLS Estimated Generalized Least Squares; a technique developed for single equations to fit a linear model using an estimated error covariance matrix.

endogenous variables Variables generated by the system of equations, stochastic.
exogenous variables Variables generated outside the system of equations, nonstochastic.

F.I. Fit Index

GLS Generalized Least Squares; a technique developed for single equations to fit a linear model using a known error covariance matrix.

heteroskedasticity Variances of the error terms are not equal across the range of the sample data; error terms are not identically distributed.

homoskedasticity Variances of the error terms are equal across the range of the sample data; error terms are identically distributed.

I3SLS Iterated Three Stage Least Squares; the estimated error covariance matrix from the first 3SLS fit is used to fit the system of equations again. The process is repeated until convergence occurs.

iid independent and identically distributed

IMLSLS Iterated Multistage Least Squares; the estimated error covariance matrix from the first MSLS fit is used to fit the system of equations again. The process is repeated until convergence occurs.

inconsistent The probability that the estimate is within a small deviation from the population value does not approach one as the number of samples is increased.

LHS Left Hand Side

M.A.D. Mean Absolute Deviation

M.D. Mean Deviation

MLE Maximum Likelihood Estimator

MSLS Multistage Least Squares; derived in this thesis for fitting systems of equations which are simultaneous and contemporaneously correlated and have with non-iid error terms.

non-iid Either not independent or not identically distributed or both.

OLS Ordinary Least Squares; techniques include simple linear regression, multiple linear regression, nonlinear least squares, and regression of weighted models. One of the assumptions for OLS is that the error terms are iid.
**psps** permanent sample plots; plots that are established and marked so that future measurements from the same trees can be taken.

**RHS** Right Hand Side

**serial correlation** Dependence between the current error term and the previous one(s).

**simultaneity bias** The bias in the estimated coefficients which results if an equation with endogenous variables on the RHS is fit using OLS, GLS, or EGLS.

**simultaneous equations** A system of equations wherein the LHS variable of one or more equations also appears on the RHS of one or more equations in the system.

**SUR** Seemingly Unrelated Regression; a technique to fit systems of equations which are contemporaneously correlated but not simultaneous.