

**GROWTH PREDICTION OF RECENT PERMANENT SAMPLE PLOTS
FOR FOREST INVENTORY PROJECTION**

By

Guillaume Thérien

B. Sc. App. (Foresterie) Université Laval

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

in

THE FACULTY OF GRADUATE STUDIES
FORESTRY

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1990

© Guillaume Thérien, 1990

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 June 26 1990

Abstract

Permanent sample plots have become the main source of information for estimating models which quantify the dynamic processes of a forest. Fitted models allow for projecting inventories, used to determine timber production and many forest management decisions. The quality of these models is largely dependent on the quality of the information provided by the permanent sample plots. However, the pool of information contained in recent permanent sample plots is limited. Efficient estimation techniques must use all the information available from such plots.

Current estimation techniques can be improved. Existing techniques employed in forestry have failed to recognize the random nature of the individual model characterizing each plot. On the other hand, techniques designed for remeasured entities in other scientific fields do not address particular forestry situations such as the small number of remeasurements or the irregularity of remeasurements. A framework for estimating forestry growth models which recognizes the individuality of each plot and special forestry situations is presented in this dissertation.

The proposed framework is a two-stage estimation technique, in which the growth rate of a permanent sample plot is considered analogous to the interest rate on a bank account. The first stage estimates the growth rate after removing the time effect. The second stage, based on Von Bertalanffy's growth curve, relates growth rate to site index and volume at the beginning of the growing season. The proposed predictor of future growth rates, the "weighted predictor," is a weighted average between the growth rate observed on a plot and the growth rate predicted from the second-stage model. The

weighted predictor is then used to compound the current volume of a plot. An estimate of the variance of the prediction can also be computed.

Table of Contents

Abstract	ii
List of Tables	vii
List of Figures	viii
Acknowledgement	ix
1 Introduction	1
2 Literature Review	7
2.1 Forestry Growth and Yield Literature	7
2.2 The Econometrics Literature	14
2.2.1 Estimated Generalized Least Squares	14
2.2.2 All Parameters Are Constant	16
2.2.3 Only the Intercepts Vary	17
2.2.3.1 The General Model	17
2.2.3.2 The Intercepts Are Fixed	18
2.2.3.3 The Intercepts Are Random	19
2.2.3.4 Recapitulation	19
2.2.4 All Parameters Vary	20
2.2.4.1 The General Model	20
2.2.4.2 The Parameters Are Fixed	21
2.2.4.3 The Parameters Are Random	22

2.2.4.4	Recapitulation	25
2.2.5	Summary of the Econometrics Literature	26
2.3	The Growth Curve Literature	26
2.4	Conditional Densities and Expectations	28
2.5	Overview	31
3	Statistical Methods	33
3.1	The Timber Capital	33
3.1.1	The Interest Rate Analogy	34
3.1.2	Uncertainty and Growth Rate	37
3.2	The Statistical Model	38
3.2.1	Notational Conventions	39
3.2.2	First-Stage Model	40
3.2.2.1	Model Definition	40
3.2.2.2	Error Components	42
3.2.2.3	Assumptions	43
3.2.2.4	Matrix Notation	47
3.2.3	Second-Stage Model	49
3.2.3.1	Background	50
3.2.3.2	Model Definition	52
3.2.3.3	Assumptions	53
3.2.3.4	Matrix Notation	54
3.2.4	Combined Model	55
3.2.4.1	Assumption	55
3.2.4.2	Matrix Notation	55
3.2.5	Summary	56

3.3	Parameter Estimation	56
3.3.1	Estimation When Covariance Matrices Are Known	57
3.3.1.1	Parameter Estimators	57
3.3.1.2	Estimation of the Variance of the Predictor	61
3.3.2	Covariance Matrix Estimators	64
3.3.2.1	First-Stage Estimators	64
3.3.2.2	Second-Stage Estimators	67
3.3.3	The Best Linear Unbiased Predictor and Its Variance	69
3.4	Model Predictions	69
3.4.1	Volume Predictor	70
3.4.2	Confidence Intervals Estimator	72
4	Numerical Example	75
4.1	Description of the Data Base	75
4.2	Goodness of Fit Criteria	79
4.3	Results	81
5	Discussion	89
5.1	The Basic Concept	89
5.2	The Statistical Model	91
5.3	Results	96
6	Conclusions	99
	References Cited	102

List of Tables

4.1	Time Intervals	77
4.2	Frequency of Volumes by Initial Age and Projection Time Class	78
4.3	Frequency of Volumes by Initial Age and Total Lag Class	79
4.4	Estimation of Observed Growth Rates	82
4.5	Overall Results from the First- and the Second-Stage Estimators and from the Weighted Predictor	84
4.6	Bias and (Relative Bias) by Initial Age and Projection Time Class	86
4.7	Bias and (Relative Bias) by Initial Age and Total Lag Class	86
4.8	Root Mean Squared Error and (Relative Root Mean Squared Error) by Initial Age and Projection Time Class	87
4.9	Root Mean Squared Error and (Relative Root Mean Squared Error) by Initial Age and Total Lag Class	87
4.10	Mean Absolute Deviation and (Relative Mean Absolute Deviation) by Ini- tial Age and Projection Time Class	88
4.11	Mean Absolute Deviation and (Relative Mean Absolute Deviation) by Ini- tial Age and Total Lag Class	88

List of Figures

3.1	Growth Rate over Time for a Sigmoidal Yield Curve	41
4.2	Distribution of Site Index	76
4.3	Individual Plot Volume versus Age	76

Acknowledgement

I would like to thank my supervisor Dr. A. Kozak as well as the rest of my supervisory committee, Dr. P. Marshall, Dr. K. Iles, Dr. P. de Jong, and Dr. D. Tait for their direction and guidance. Their contributions to my thesis and forest biometrics education were enormous, as was their patience. I extend my appreciation to Dr. N. E. Heckman and Dr. J. G. Cragg for their short but invaluable contributions to my work.

Funding for this research was supplied by the University of British Columbia through numerous scholarships, fellowships, teaching assistantships, research assistantships, *etc.*

Special thanks to MacMillan Bloedel Ltd. for allowing me access to the best data base in the country.

Finally, I would like to thank my peers of the first hour: Mrs Karen Jahraus, Dr. Margaret Penner, Dr. Valerie LeMay, and Dr. Jim Thrower, for their constant encouragement and support. They initiated an instructive forum where ideas and opinions were debated freely. Many thanks to Valerie who also commented on zillions of earlier drafts, and to Margaret whose own acknowledgement "inspired" this one.

Chapter 1

Introduction

The main purpose of forest management is to maintain or increase the yield of the different forest resources. This goal requires a knowledge of current inventory and growth potential. Growth potential is the change in the current inventory that would occur without harvesting. However, estimating growth potential is a difficult task because the forest is a dynamic system.

Quantifying the dynamic processes in a forest is essential for forecasting the future inventory. One way of quantifying the timber aspect of the dynamic processes of a forest is to use permanent sample plots. Monitoring the same piece of land over time represents a logical way to assess timber growth. In British Columbia, permanent sample plots were first established for growth and yield purposes in the late twenties and the early thirties. They became more popular in the sixties and seventies because of an increased interest in predicting yield from managed stands (Marshall and Jahraus 1987). Today, many agencies have a network of recent permanent sample plots, which have had only a few measurements.

Timber dynamics can be represented by a mathematical model that ideally has the following properties: representativity, tractability and simplicity. This model is called a growth model. The model should be representative of the biological behaviour and of the stochastic nature of this behaviour; it should be tractable so that estimation is possible, and simple in its parameterization.

Modeling growth of permanent sample plots allows for the projection of sample plot inventories. Projections of temporary sample plot inventories can also be obtained. This information will be needed to estimate the timber production of a management unit and to determine an appropriate level of harvesting for this management unit. Most techniques used today to quantify inventory changes do not take advantage of all the information available from permanent sample plots. Techniques especially designed for remeasured entities are required. As most forest agencies have a large proportion of recent permanent sample plots, these techniques must be compatible with sample plots that have had only a few measurements.

The objective of this dissertation is to define a modeling framework that will make use of all the information available from recent permanent sample plots in order to project an inventory with precision and accuracy. Estimates of the variability of the projections is also of special interest. For simplicity, only linear models are considered.

The general linear statistical model of permanent sample plot collections can be expressed by:

$$Y_{ij} = \beta_{0ij} + \beta_{1ij}X_{1ij} + \dots + \beta_{Kij}X_{Kij} + e_{ij}$$

where Y_{ij} is the dependent variable measured on individual i ($i = 1, \dots, N$) at time j ($j = 1, \dots, T$); X_{pij} is the variable p ($p = 1, \dots, K$) measured on individual i at time j ; β_{pij} is the parameter associated with X_{pij} , β_{0ij} is the intercept, and e_{ij} is the error term on individual i at time j .

Judge *et al.* (1985, p. 515–516) classified the parameter estimation techniques of such models in 5 broad categories:

1. all parameters are constant over individuals and time;
2. the intercept varies over individuals only, the other parameters are constant over individuals and time;

3. the intercept varies over individuals and time, the other parameters are constant over individuals and time;
4. all parameters vary over individuals only;
5. all parameters vary over individuals and time.

In forestry, most techniques developed for fitting models on permanent sample plots have assumed constant parameters over time and individuals. This fails to recognize the individuality of each sample plot. Other scientific fields dealing with collections of remeasured individuals have used models with at least one parameter varying over individuals. These models can “individualize” the general model. However, recent permanent sample plots have some particular problems that models used in other fields do not address. These include:

1. the plots are not the same age at the first measurement;
2. the time interval between measurements is not always the same within a plot;
3. plots are not remeasured at the same time;
4. the number of remeasurements are small.

As a general rule, special attention must be given to the error structure of the model and the assumptions about the covariance matrix of the random terms if efficient estimators are required. The usual assumptions required for simple or multiple linear regression are frequently not valid with remeasured individuals. Because each plot is remeasured at different times, the plots are often referred to as a time-series. Observations from a time-series are often correlated; this type of correlation is called serial correlation. Observations measured at the same time can also be correlated; this type of correlation is

called contemporaneous correlation. Both types of correlation provide information about the remeasured entities needed to estimate model parameters efficiently.

In this dissertation, a new framework for projecting recent permanent sample plot inventories is proposed. This new framework addresses the complexities of remeasured entities particular to forestry. The variable of interest is growth rate, that is, the change in volume per unit of initial volume.¹ For short projections, an intuitive technique to predict future volumes on a given plot is to compound the current volume with the growth rate recently observed on this plot. Another technique, requiring a collection of permanent sample plots, is to regress the observed growth rates on stand attributes. The volume is then updated with the growth rate predicted from the regression. Another option, the one favoured in this thesis, is to use an average of both techniques, with each technique's growth rate estimates weighted proportionally to its reliability.

The framework employed a two-stage model. In the first stage, average recent growth rates are obtained for individual plots. In the second stage, model parameters are estimated assuming growth rates to be a function of site index and initial volume. The predictor for a sample plot is a weighted average between the growth rate estimated in the first stage and the growth rate predicted from the second-stage model.

Statistically, the framework can be explained in the following way. A general model exists, explaining the growth rate from two stand attributes: site index and initial volume. However, each individual permanent sample plot follows its own model; these individual models vary randomly around the general model. The individual model can be expressed as the general model plus an error term specific to a particular plot. The actual growth

¹This variable is also sometimes referred to as relative growth rate. In this thesis the term growth rate will be used.

rate observations also vary, but around the individual model. The framework can therefore be considered a hierarchical model: the observations are random variations of the general model's random variation.

Repeated observations on the same individual and an estimate of the general model can yield a prediction of the random error term specific to a particular plot. Adding this term to the estimated general model gives a prediction of the individual model. The individual model is "predicted" and not "estimated" because the individual model is a random variation of the general model. The predicted individual model is then used to provide a prediction of the growth rate used to update the current volume.

There are many reasons to prefer an estimate of the individual growth rate rather than a global estimate. The individual estimate can be used as a decision aid for plot remeasurement scheduling. It can also be employed in sampling with partial replacement to update the plots that have not been remeasured. A third possible use of individual growth rates is the ordination of the sample plots in groups to study the causes of the differences among groups.

Most techniques used in forestry have failed to recognize the stochastic nature of this individual relationship. They also have failed to take into account that many permanent sample plots in North America are less than 25 years of age and have been measured less than 5 times. The proposed framework will provide a better understanding of the information given by permanent sample plots, especially recent ones, and will result in more efficient fitting techniques.

The dissertation has been organized according to the following format: a brief presentation of the past work in growth and yield and in other fields using data sets containing remeasured individuals is discussed in Chapter 2; the statistical model, the various assumptions and the estimators are discussed in Chapter 3; a numerical example and results

are presented in Chapter 4; these results, along with the statistical model, are discussed in Chapter 5, and, finally, conclusions are drawn in the last chapter.

Chapter 2

Literature Review

Permanent sample plot data bases are a collection of plots remeasured a number of times. Remeasured data bases are not unique to forestry. Research scientists in business, economics, biology and medicine have developed techniques customized to their specific problems involving entities with repeated measurements over time.

An overview of these techniques is presented in the following pages. A brief history of forestry growth and yield can be found in the first section. The second section groups the techniques developed in econometrics. Much attention has been given in business and economics research to data bases similar to permanent sample plot data bases. Research on biological growth curves is reviewed in the third section. Because trees are biological organisms, general knowledge of biological growth can help understand tree or stand growth. Finally, the last section is a brief presentation of conditional densities and expectations. Useful matrix inversion rules are also given. These basic rules will help in understanding the statistical model suggested in Chapter 3 to project inventory.

2.1 Forestry Growth and Yield Literature

Predicting yield, along with determining the current inventory, is a common forest management activity. Availability of reliable information, computing technology and information requirements have influenced the quality of the information desired from forest inventory projections. Permanent sample plot data bases have become the main source

of information for these projections. A historical perspective of the uses of permanent sample plot data as a source of information is presented in the following pages.

Yield is highly correlated with a few well-known variables: species, age, ecological association, climate, edaphic conditions, among others. Nineteenth century German foresters used the variables age, species, and site class to create their normal yield curves. "Normal" in this case merely meant that the variability in stocking was omitted by using only fully-stocked, even-aged, monospecific temporary sample plots. At any point in time, these curves represented the upper bound or potential yield of a stand. The interpretation of predictions from these tables was left to the manager's judgment and experience.

To make these tables applicable to understocked forests, Gerhardt (1930, in Davis and Johnson 1987; p. 106) developed a simple correction factor. Growth was estimated from the percentage of stocking compared to that of a fully-stocked forest, the predicted growth for a fully-stocked forest and the species. With the correction factor, it was assumed that all stands were moving toward full stocking at a rate depending on the species and the stocking.

Schumacher (1939) published a simple empirical yield function where the logarithm of volume was a function of site index (S) and the reciprocal of age ($1/A$):

$$\ln V = \beta_0 + \beta_1(1/A) + \beta_2(S) + \beta_3(S/A)$$

No statistical assumptions were made to develop this model; no statistical inference could be drawn from this model. The parameters were fitted using least squared error as the criterion of optimality. Schumacher's model was among the first of the mathematical expressions for yield.

Buckman (1962) recognized that serially correlated data from permanent plots were providing misleading significance levels and confidence intervals when analyzed as independent observations.

Clutter (1963) recognized the need for compatible growth and yield models, where predicted growth is the derivative of predicted yield. He justified the use of volume on a logarithmic scale to fit his growth and yield equations by assuming a better compatibility in general with "the statistical assumptions customarily made in regression analysis (linearity, normality, additivity and homogeneity of variance)."

Leak (1966) developed a technique for providing a large-sample estimator of the overall slope of a random coefficient model. He first assumed a linear relationship between the dependent and the independent variable for each sample plot, with the error terms independently distributed as $N(0, \sigma_i^2)$, where σ_i^2 is the variance of an error term observed on sample plot i . These individual regressions were solved using the least squares technique. On each plot, he assumed the same number of observations and a similar distribution of the independent variable, X , around the average, \bar{X} (i.e., the corrected sums of squares of the variable X were approximately equal). The expected value of any individual slope was the overall slope. This was estimated as the simple average of the individual slopes. Assuming that σ_i^2 was equal among plots, the standard error of the estimated overall slope was the standard error of the mean slope.

Curtis (1967) developed a method to estimate the gross yield of Douglas-fir. He recognized that the usual assumption of independence was probably violated when fitting equations to permanent sample plots. To handle the problem, he first fitted linear growth equations on each plot without regard to autocorrelation using the least squares technique. Then, the coefficient of determination was computed for u_{i-1} and u_i , the consecutive estimated residuals on a plot. If the coefficient of determination r^2 was small, autocorrelation was considered negligible; otherwise, the analysis was repeated using a single randomly selected point per plot.

Swindel (1968) showed the theoretical bounds on the bias of the variance of parameter estimates when serially correlated data were assumed independent for multiple linear

models estimated with least squares. Sullivan and Reynolds (1976), using Swindel's work, provided an example for data with two measurements.

Sullivan and Clutter (1972) developed growth and yield models using permanent sample plots with two measurements per plot. They assumed common parameters over plots and time, homogeneous variance within a measurement period, serial correlation ρ between the first and the second measurements, and no contemporaneous correlation, in that the volumes measured in a given year were uncorrelated. If $\mathbf{V}'_i = (V_{i1} \ V_{i2})$ is the vector containing the first and second measurements of volume on plot i , then, for any $i \neq j$, \mathbf{V}_i and \mathbf{V}_j are uncorrelated and

$$\text{cov}(\mathbf{V}_i) = \text{cov}(\mathbf{V}_j) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

where ρ is the correlation between the first and the second measurement. They derived the maximum likelihood estimates of multiple linear models by an iterative procedure. Seegrist and Arner (1978) and Arner and Seegrist (1979, 1980) extended this method to allow multiple and varying number of remeasurements.¹

Ferguson and Leech (1978), fitted a two-stage model drawing on the theory relating to random coefficient modeling. In the first stage, they fitted a quadratic model to each plot with the reciprocal of age as the independent variable using ordinary least squares. They assumed that the coefficients were different among plots:

$$\ln V_{ij} = \beta_{i0} + \beta_{i1}(1/A_{ij}) + \beta_{i2}(1/A_{ij})^2 + e_{ij}$$

where V_{ij} is the volume measured on plot i at time j , β_{ik} is the coefficient k on plot i , A_{ij} is the age of plot i at time j , and e_{ij} is an error term. Using matrix notation, \mathbf{V} is the column-vector of the logarithmic volumes; β is the vector of parameters to be estimated; \mathbf{X} is the matrix of independent variables, and \mathbf{e} is the vector of error terms. The first

¹In econometrics nomenclature, this type of estimation is called classical pooling.

stage can be written

$$\mathbf{V} = \mathbf{X}\beta + \mathbf{e}.$$

With the assumptions that \mathbf{e} is distributed with mean vector zero and variance $\mathbf{\Omega}$ and is uncorrelated with β or \mathbf{X} , the least squares estimate of β , $\hat{\beta}$, is unbiased with variance $[\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1}$.

They then considered β as a random variable that could be predicted from certain independent variables (second-stage models). The set of independent variables to predict β do not need to be related to the set of independent variables used to predict \mathbf{V} . They fitted three different models: to predict the intercept, the coefficient of the linear term and the coefficient of the quadratic term. The second-stage model structure was

$$\beta_{ik} = \sum_l \theta_{kl} Z_{ikl} + u_{ik}$$

where θ_{kl} is the parameter associated with the variable l for the coefficient k ; Z_{ikl} is the corresponding explanatory variable on plot i , and u_{ik} is the error term. Using matrix notation, the second stage can be written

$$\beta = \mathbf{Z}\theta + \mathbf{u}.$$

Assuming that \mathbf{u} is distributed with mean vector zero and variance $\mathbf{\Sigma}$ and is uncorrelated with θ and \mathbf{Z} , the random variable β is distributed with mean vector $\mathbf{Z}\theta$ and variance $\mathbf{\Sigma}$.

Because β was considered a random variable, the distribution of $\hat{\beta}$ in the first stage was actually the conditional distribution of $\hat{\beta}$ given β with

$$\begin{aligned} E[\hat{\beta}|\beta] &= \beta, \quad \text{and} \\ \text{var}(\hat{\beta}|\beta) &= [\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1} \end{aligned}$$

Combining the conditional distribution of $\hat{\beta}$ given β with the distribution of β , the unconditional distribution of $\hat{\beta}$ will have

$$E[\hat{\beta}] = \mathbf{Z}\theta, \quad \text{and}$$

$$\text{var}(\hat{\beta}) = [\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}]^{-1} + \mathbf{\Sigma}.$$

Because $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are unknown, they must be replaced by estimates. In the first-stage model, Ferguson and Leech assumed homogeneous variance within plots, no serial correlation, and heterogeneous variance among plots. In the second-stage, they assumed a homogeneous covariance matrix for the coefficients and independence between sets of coefficients measured on different plots. With these assumptions they developed a feasible generalized least squares (FGLS) algorithm. They also investigated more restrictive assumptions about the error structure. Commenting on this study, Davis and West (1981) claimed that no proper confidence intervals nor test of significance can be computed with this approach. Ferguson and Leech (1981) replied that asymptotically normal statistics have proved adequate for most practical purposes using large samples.

Garcia (1983) fitted a height growth model with a stochastic differential equation. Height growth was assumed to be a function of height (H):

$$\partial H / \partial t = f(H)$$

The function of height was assumed to be $\phi_1 H^{\phi_2} - \phi_3 H$, where ϕ_1 , ϕ_2 , and ϕ_3 are the function parameters. This function integrates to the well-known Chapman-Richards function (Chapman 1961, Richards 1959). The model included an error term to account for the random environmental variation in growth and the measurement error. The environmental error was assumed to be a Wiener process. This means that the variation in H accumulated over a short time interval is normally distributed with zero mean and a variance increasing with the interval length, and that errors for nonoverlapping time intervals are independent. The measurement error was assumed to be a function of observed height. The plots were assumed statistically independent. The parameters were simultaneously estimated with an iterative maximum likelihood procedure. Some parameters were global, that is identical across plots; some were local, that is site specific.

Gregoire (1987) applied error component modeling to permanent sample plot data. The basic model was

$$Y_{ij} = \sum_p \beta_p X_{pij} + u_{ij}$$

$$u_{ij} = e_i + r_j + v_{ij}$$

with X_{1ij} defined as 1. The error was divided into three components: one for plot-dependent variation (e_i), one for time-dependent variation (r_j) and one, (v_{ij}), for the variation unexplained by plot-dependent or time-dependent effects; Y_{ij} was the dependent variable measured on plot i at time j , and X_{pij} was the variable p measured on plot i at time j . The expected value of each component of the error term was assumed to be zero. The error term components were also assumed to be uncorrelated with the independent variables and uncorrelated with each other. Alternative error covariance structures were analyzed. When compared to the least squares technique, with the undivided error terms assumed independently and identically distributed, the results were not conclusive. In general, the error component models performed better than the least squares for the likelihood criterion but worse for the minimum prediction error criterion. Gregoire also pointed out the need to find good statistics to compare alternative models.

Biging (1985) derived a site index model using random coefficient modeling theory in order to account for between-tree differences in individual tree height growth. The error structure was identical to Swamy's (1970) model, and the parameters were estimated using Maddala's (1977) estimators.

Finally, Tait *et al.* (1988) developed a simple, two-equation growth model where the growth of the average tree was a function of the average tree volume, site quality and the number of trees per hectare, and the mortality was a function of an index of the rate of area expansion of a tree. They simultaneously estimated the best model parameters (*i.e.*, the global parameters) with the best initial values for each plot (the local parameters).

This type of analysis is close to an analysis of covariance where each plot is considered a different treatment with two covariates, the initial stand density and volume.

This brief historical report illustrates the change of attitude toward the information pool contained in permanent sample plot data bases. It is now recognized that an observation at a point in time on a given individual provides partial information, on one hand, on the same individual at other points in time and, on the other hand, on other individuals at the same point in time. Efficient parameter estimation techniques must exploit this information.

2.2 The Econometrics Literature

In econometrics literature, a data base that provides repeated measurement for each of a number of entities is called a pooled cross-sectional and time-series data base (Dielman 1983). The term panel data is also often used. This type of data base is very common in econometrics. The abundance of information has led econometricians to develop techniques that take advantage of this information to explain and predict economic phenomena. Because of the similarities with forestry permanent sample plots, there has been a growing interest in forestry to adapt techniques developed in econometrics (*e.g.*, Furnival and Wilson 1971, Ferguson and Leech 1978, Gregoire 1987, LeMay 1988). The literature about pooled cross-sectional and time-series data is extensive and only a brief overview of the major developments will be presented here.

2.2.1 Estimated Generalized Least Squares

When a simple or a multiple linear model is fitted to predict or to explain the outcome of a dependent variable, the most common assumptions about the error term are zero expectation and spherical disturbance, (*i.e.*, the covariance matrix of the error term is

$\sigma^2\mathbf{I}$). These models are easily estimated with the ordinary least squares (OLS) technique. However, the assumption of spherical disturbance is often not met. Another technique, called *estimated generalized least squares*, allows for estimating efficiently the parameters of such models.

The general linear model can be expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$$

where \mathbf{Y} is the column-vector of dependent variables, \mathbf{X} is a matrix of independent variables, β is the column-vector of model parameters, and \mathbf{u} is the column-vector of random error terms. Aitken (1935) showed that if the covariance matrix of \mathbf{u} is Ω , then the best linear unbiased estimate (BLUE) of the parameters is

$$\hat{\beta} = [\mathbf{X}'\Omega^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}.$$

This estimate has been called the generalized least squares (GLS) estimate or Aitken estimate. If the error terms are assumed to be normally distributed, the GLS estimate is also the maximum likelihood estimate.

Unfortunately, most of the time, Ω is unknown. In this case, a two-step technique is required to obtain what has been called the estimated generalized least square (EGLS) estimate or the feasible generalized least square (FGLS) estimate. The first step of the EGLS technique is to estimate Ω . In the most general case, Ω will contain $M(M+1)/2$ unknowns where M is the total number of observations. Because the number of unknowns is larger than the number of observations, assumptions about the structure of Ω must be made to restrict the number of unknown elements to a manageable number. In the second step, $\hat{\Omega}$, the estimate of Ω , is simply substituted into the Aitken estimate. Consequently, the FGLS estimate of β is

$$\hat{\hat{\beta}} = [\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{Y}$$

which is the same as the GLS estimate except for Ω which was replaced by an estimate. The double hat above β emphasizes that it is estimated from a two-step technique.

If \mathbf{Y} is replaced by $\mathbf{X}\beta + \mathbf{u}$ in the EGLS estimate, then it is easily shown that

$$\hat{\hat{\beta}} = \beta + (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{u}.$$

Because $\hat{\Omega}$ and \mathbf{u} are usually correlated, inferences about β need to be based on the asymptotic distribution of $\hat{\hat{\beta}}$ (Judge *et al.* 1985; p. 175). Generally, $\hat{\hat{\beta}}$ will be asymptotically normally distributed if \mathbf{u} is normally distributed and $\hat{\Omega}$ is a consistent estimator for Ω . An estimator $\hat{\alpha}$ is a consistent estimator for α if, for any $\zeta > 0$,

$$\lim_{M \rightarrow \infty} P[|\hat{\alpha} - \alpha| > \zeta] = 0.$$

The EGLS technique is the general approach for obtaining the BLUE of the parameters of any linear model. The OLS technique is a special case, particularly restrictive, of the EGLS technique. With pooled cross-sectional and time-series data, the assumed structure of β will have a supreme importance on the structure of the covariance matrix of the error terms. This is why the EGLS technique for pooled cross-sectional and time-series data has been organized in three different sections corresponding to the general assumptions about the structure of the model parameters: all parameters are constant, only the intercepts vary, and all parameters vary.

2.2.2 All Parameters Are Constant

Each individual measured over time can be represented by its own model

$$\mathbf{Y}_i = \mathbf{X}_i\beta_i + \mathbf{u}_i$$

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{bmatrix} \quad \mathbf{X}_i = \begin{bmatrix} 1 & X_{1i1} & \dots & X_{Ki1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1iT} & \dots & X_{KiT} \end{bmatrix} \quad \beta_i = \begin{bmatrix} \beta_{i0} \\ \vdots \\ \beta_{iK} \end{bmatrix} \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

When it is assumed that all individuals follow a similar behaviour, the model parameters are assumed to be equal over individuals. The variation between individuals is then explained by the stochastic error term only. This is called *classical pooling* (CP). If the error terms on the observations from a single individual were uncorrelated, the observations from this individual would oscillate around their expected values. However, because of serial correlation, a given individual can perform constantly above or below the expected behaviour. Classical pooling is used when there is no reason to believe that the different behaviour between individuals can be explained other than by serial correlation and chance.

2.2.3 Only the Intercepts Vary

Sometimes, it is possible to explain partially the difference between individuals by their initial states. In these cases, it is assumed that the partial derivatives, that is the slopes of each independent variable, are constant over individuals and time, and that the intercepts will vary over individual and time.

In this section, the general model describing such a situation is presented. Alternative approaches to parameter estimation depend on the assumptions about the nature of the intercepts. The intercepts can be assumed to be fixed or random. Both alternatives lead to different estimation techniques that are discussed separately. Finally, the model and the estimation techniques are summarized in a short recapitulation.

2.2.3.1 The General Model

In the most general case, the error term of the general linear model can be divided into three components: one to account for the difference between individuals, one to account for the difference between initial time periods, and a third for the variables unaccounted

for in the model that are not individual-dependent or time-dependent. This gives

$$Y_{ij} = \bar{\beta}_0 + \sum_{p=1}^K \beta_p X_{pij} + e_i + r_j + v_{ij}$$

where Y_{ij} is measurement j on plot i , $\bar{\beta}_0$ is the average intercept, X_{pij} is the variable p measured on plot i at time j , e_i is the individual error term, r_j is the temporal error term, and v_{ij} is the unexplained variation. The individual intercept is defined by combining e_i and r_j with the average intercept. These different intercepts will likely arise from omission of important unobservable or unmeasurable variables in the model (Dielman 1989, p. 49).

2.2.3.2 The Intercepts Are Fixed

In this case, the individual error component and the temporal error component are assumed to be fixed with mean zero (*i.e.*, $\sum e_i = 0$ and $\sum r_j = 0$) and the third unexplained component is random with mean zero and variance σ^2 . The estimation technique is called *analysis of covariance* (ANCOVA) or *least squares with dummy variables*. Mundlak (1961) was among the first to use the ANCOVA for analyzing pooled cross-sectional and time-series data.

There are a few disadvantages of ANCOVA. Since the different intercepts are considered fixed, they are using up as many degrees of freedom. Also, the different components of the error term represent some ignorance; the model does not explain the differences between initial states. Maddala (1971) suggested that this specific ignorance was not really different from the general ignorance ν_{it} and should be considered in a similar way. This would also decrease the number of fixed parameters to be estimated, increasing therefore the number of degrees of freedom. This led to the assumption of random intercepts.

2.2.3.3 The Intercepts Are Random

In this case, it is assumed that the various components of the error term are random effects with mean zero and specific variance:

$$E[e_i] = E[r_j] = E[v_{ij}] = 0$$

$$E[e_i^2] = \sigma_i^2$$

$$E[r_j^2] = \sigma_j^2$$

$$E[v_{ij}^2] = \sigma_{ij}^2.$$

It is also usually assumed that e_i , r_j and v_{ij} are mutually independent. The estimation technique based on these assumptions is called *error component modeling* (ECM).

Wallace and Hussain (1969) suggested estimating the variance components using the residuals obtained from the OLS estimate of the individual models. These estimators are consistent when the independent variables are assumed nonstochastic. Other estimators of the error components have been developed (*e.g.*, Amemiya 1971, Swamy and Arora 1972, Rao 1972, Fuller and Battese 1974). All these estimators have similar asymptotic properties. Consequently, the choice of an estimator cannot be based on its asymptotic properties. One must consider the simplicity of the estimator, finite sample results and Monte Carlo studies of small sample performance (Baltagi 1981). The finite sample properties of some estimators have been investigated analytically by Swamy and Arora (1972), Swamy and Mehta (1979) and Taylor (1980). Monte Carlo studies were published by Arora (1973), Maddala and Mount (1973) and Baltagi (1981). Swamy and Mehta (1973) analyzed the error component model in a Bayesian framework.

2.2.3.4 Recapitulation

The difference between individuals can sometimes be explained by different initial states (or model intercepts). These different intercepts come from partitioning the error term

into generally three components: an individual error term, a temporal error term, and an unexplained error term, and from the combination of the common intercept with the individual and the temporal error terms. If the individual and the temporal error terms are assumed to be fixed, then the ANCOVA is the appropriate EGLS technique. If the individual and temporal error terms are assumed to be random, then ECM must be used.

2.2.4 All Parameters Vary

Some time-series observations of different individuals are highly correlated with the same set of independent variables but with slopes varying over individuals. Not only are their initial states different but also the increment in the dependent variable for one unit of change in one of the independent variables. This will occur if an important variable for all the individuals is omitted from the model or if a variable important for some individuals is omitted.

This section follows the same format as the previous section. The general model describing the situation is presented first. The parameter estimation technique when the parameters are assumed to be fixed comes next, followed by the technique for random parameters. Last, the model and the techniques are briefly recapitulated.

2.2.4.1 The General Model

The individual model is

$$\mathbf{Y}_i = \mathbf{X}_i \beta_i + \mathbf{u}_i$$

where

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{bmatrix} \quad \mathbf{X}_i = \begin{bmatrix} 1 & X_{1i1} & \dots & X_{Ki1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1iT} & \dots & X_{KiT} \end{bmatrix} \quad \beta_i = \begin{bmatrix} \beta_{i0} \\ \vdots \\ \beta_{iK} \end{bmatrix} \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

Then, the full model can be written:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & & \\ & \ddots & \\ & & \mathbf{X}_N \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}.$$

2.2.4.2 The Parameters Are Fixed

With fixed parameters, efficient parameter estimates can be obtained using OLS if there is no contemporaneous correlation. When contemporaneous correlation is present, Zellner (1962) developed a technique using this supplementary information. He called the technique *seemingly unrelated regressions* (SUR). Estimators obtained with this technique are more efficient than simply applying least squares to each individual independently.

The main assumptions about the error structure are: $E[\mathbf{u}_{ij}] = 0$, $E[\mathbf{u}_i \mathbf{u}_l'] = \sigma_{il} \mathbf{I}_T$, and \mathbf{X}_i is a fixed matrix in repeated samples. N is the number of individuals, K is the number of independent variables in the model, and T is the number of observations on each individual. Based on the assumptions about the error terms, the covariance matrix of the error terms can be written

$$\Omega = \Sigma \otimes \mathbf{I}_T$$

where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \sigma_{NN} \end{bmatrix}$$

When the contemporaneous correlation between individuals i and l is equal to zero, $\sigma_{il} = 0$ for $i \neq l$, the estimates are easily shown to be identical to the OLS estimates

computed independently on each individual. The SUR estimator will also be the same as the OLS estimator if the \mathbf{X}_i are identical for all individuals. If \mathbf{u} is normally distributed, and $\lim_{T \rightarrow \infty} [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}$ is a finite nonsingular matrix, the EGLS estimate is a minimum variance unbiased estimator, and is asymptotically efficient.

To estimate $\boldsymbol{\Omega}$, Zellner (1962) suggested using the residuals from the regression of \mathbf{Y}_i on either the independent variables in the i^{th} equation alone or all the independent variables in the system. Writing the residuals from either method as \hat{u}_{ij} , the elements of $\boldsymbol{\Sigma}$ (or $\boldsymbol{\Omega}$) are estimated by

$$\sigma_{il} = \sum_{j=1}^{j=T} \hat{u}_{ij} \hat{u}_{lj} / (T - K).$$

Zellner (1962) also suggested iterating the estimator by recomputing the residuals using $\hat{\beta}$ to get a new estimate of $\boldsymbol{\Omega}$, and so on until convergence occurred. Dhrymes (1971) showed that the iterative estimator is equivalent to the maximum likelihood estimator.

Parks (1967) and Kmenta and Gilbert (1970) extended the theory to include the case where the disturbances are both serially and contemporaneously correlated. Schmidt (1977) examined the SUR model with missing observations. Swamy and Mehta (1975) also studied SUR models with missing observations but in a Bayesian framework. Phillips (1985) derived the exact finite sample properties of the SUR estimator. Zellner (1987, pp. 240–256), Srivastava (1973), and Zellner and Vandaele (1975) analyzed the SUR model in a Bayesian framework.

2.2.4.3 The Parameters Are Random

The sets of different parameters can also be assumed to be random. The EGLS technique for such models has been called *random coefficient modelling* (RCM; Swamy 1970, 1971, 1974). Because the model parameters link the same independent variables to the dependent variable, they could very well be related. If it can be assumed that many

independent unobservable or unmeasurable factors can influence a model parameter, this model parameter can be seen as a random variable. These excluded variables can be assumed to counteract each other, so that their overall effect is negligible. Consequently, the expected value of the model parameter is the mean parameter, and its variance is the variance of the aggregate effect of these excluded variables. The parameter β_i can be rewritten

$$\beta_i = \bar{\beta} + \mathbf{e}_i$$

where

$$\bar{\beta} = \begin{bmatrix} \bar{\beta}_0 \\ \vdots \\ \bar{\beta}_K \end{bmatrix} \quad \text{and} \quad \mathbf{e}_i = \begin{bmatrix} e_{i0} \\ \vdots \\ e_{iK} \end{bmatrix}.$$

Replacing this new expression for β_i in the original individual model yields

$$\mathbf{Y}_i = \mathbf{X}_i(\bar{\beta} + \mathbf{e}_i) + \mathbf{u}_i = \mathbf{X}_i\bar{\beta} + \mathbf{X}_i\mathbf{e}_i + \mathbf{u}_i = \mathbf{X}_i\bar{\beta} + \mathbf{w}_i$$

The error term \mathbf{w}_i is then the sum of two components, $\mathbf{X}_i\mathbf{e}_i$ and \mathbf{u}_i . The full model can be written

$$\mathbf{Y} = \mathbf{X}\bar{\beta} + \mathbf{w}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} (\mathbf{X}_1\mathbf{e}_1 + \mathbf{u}_1) \\ \vdots \\ (\mathbf{X}_N\mathbf{e}_N + \mathbf{u}_N) \end{bmatrix}.$$

Swamy (1970, 1971) made the following assumptions: the number of individuals and the number of observations on each individual are larger than the number of independent variables; the \mathbf{u}_i are independently and identically distributed with $E[\mathbf{u}_{ij}] = 0$ and $E[\mathbf{u}_i\mathbf{u}_i'] = \sigma_{ii}\mathbf{I}_T$; the \mathbf{e}_i are independently and identically distributed with $E[\mathbf{e}_{ip}] = 0$ and

$E[\mathbf{e}_i \mathbf{e}_i'] = \Delta$, and the \mathbf{u}_i and β_i are independent for every i and l . The covariance matrix of \mathbf{w} , Ω , is the block diagonal matrix with

$$\Omega_{ii} = \{\mathbf{X}_i \Delta \mathbf{X}_i' + \sigma_{ii} \mathbf{I}_T\}.$$

The GLS estimate for $\bar{\beta}$ is

$$\hat{\bar{\beta}} = [\mathbf{X}' \Omega^{-1} \mathbf{X}]^{-1} \mathbf{X}' \Omega^{-1} \mathbf{Y} = \left[\sum_i \mathbf{X}_i' \Omega_{ii}^{-1} \mathbf{X}_i \right]^{-1} \sum_i \mathbf{X}_i' \Omega_{ii}^{-1} \mathbf{Y}_i = \sum_i \mathbf{W}_i \mathbf{b}_i$$

where

$$\begin{aligned} \mathbf{b}_i &= [\mathbf{X}_i' \mathbf{X}_i]^{-1} \mathbf{X}_i' \mathbf{Y}_i \\ \mathbf{W}_i &= \left\{ \sum_i [\Delta + \sigma_{ii} (\mathbf{X}_i' \mathbf{X}_i)^{-1}]^{-1} \right\}^{-1} [\Delta + \sigma_{ii} (\mathbf{X}_i' \mathbf{X}_i)^{-1}]^{-1}. \end{aligned}$$

The GLS estimate can be seen as a weighted average of the OLS estimate applied to each individual model. Mundlak (1978) noted that $\hat{\bar{\beta}}$ can be written as a weighted average of a “between estimator” and a “within estimator.” Rao (1982) gave the conditions (unrestrictive) under which $\hat{\bar{\beta}}$ is an unbiased estimator of β .

In general, β_i is the parameter of interest. Lee and Griffiths (1979, in Judge *et. al.* 1985; p. 541) showed that the best linear unbiased predictor (BLUP) of β_i is

$$\tilde{\beta}_i = [\Delta^{-1} + \sigma_{ii}^{-1} (\mathbf{X}_i' \mathbf{X}_i)]^{-1} \sigma_{ii}^{-1} (\mathbf{X}_i' \mathbf{X}_i) \mathbf{b}_i + [\Delta^{-1} + \sigma_{ii}^{-1} (\mathbf{X}_i' \mathbf{X}_i)]^{-1} \Delta^{-1} \hat{\bar{\beta}}.$$

Note that the term “predictor” was used instead of “estimator” because β_i is a random variable. The BLUP for β_i is a weighted average of the OLS estimate of β_i and the overall estimate of $\bar{\beta}$ with the weights inversely proportional to their respective variances. In a Bayesian analysis of a random coefficient model, Smith (1973) found a numerically identical estimator. As would be expected, the simple average of the individual BLUP’s gives the overall BLUE, that is $(\sum_1^N \tilde{\beta}_i)/N = \hat{\bar{\beta}}$.

As usual, the unknown Δ and σ_{ii} must be replaced by consistent estimators to obtain the FGLS estimate. If \hat{u}_i is the OLS residual for individual i , Swamy (1970) suggested using:

$$\hat{\sigma}_{ii} = \frac{\hat{u}_i' \hat{u}_i}{T - K}$$

$$\hat{\Delta} = \frac{1}{N - 1} \left[\sum_1^N \mathbf{b}_i \mathbf{b}_i' - \frac{1}{N} \sum_1^N \mathbf{b}_i \mathbf{b}_i' \right] - \frac{1}{N} \sum_1^N \hat{\sigma}_{ii} (\mathbf{X}_i' \mathbf{X}_i)^{-1}$$

With these estimates, Swamy showed that $\hat{\beta}$ is asymptotically efficient as $T \rightarrow \infty$. Carter and Yang (1986) showed that this result was also valid when $N \rightarrow \infty$ or $T \rightarrow \infty$ if $\sigma_{ii} = \sigma^2$ (i.e., if all the variances of the disturbance terms are equal).

Amemiya (1978) investigated the case where the individual parameters are a linear function of some explanatory variables

$$\beta_i = \mathbf{Z}_i \theta + \mathbf{e}_i.$$

Swamy (1974) extended the theory to include the case with contemporaneous correlation and serial correlation.

Finally, there are also techniques to estimate models with different sets of parameters varying across individuals *and* time. These models were not considered relevant to this dissertation and are not discussed (for information concerning these models, see Judge *et. al.* 1985; p. 545–550).

2.2.4.4 Recapitulation

The difference between individuals can sometimes be explained by different sets of model parameters. These parameters can be fixed in which case the SUR technique is the appropriate EGLS technique if contemporaneous correlation is present. If the parameters

are fixed, and there is no contemporaneous correlation, then the OLS technique is appropriate. The parameters can also be assumed to vary randomly around a common average. In this case, RCM is the appropriate EGLS technique.

2.2.5 Summary of the Econometrics Literature

The best, linear, unbiased estimate (BLUE) of the parameters of a linear model is given in its general form by the estimated generalized least squares (EGLS) estimate. When the error terms are assumed to be independently and identically distributed, the EGLS estimate is equivalent to the ordinary least squares (OLS) estimate. However, this rarely is the case with models fitted on pooled cross-sectional and time-series data. The structure of the covariance matrix of the error term is highly influenced by the assumption about the individual model parameters. There are three general cases: the individual models share the same parameters, the individual models have different intercepts but common slopes, or the individual models have different sets of parameters altogether. These different parameters can be assumed to be fixed or random. Classical pooling (CP) is the appropriate technique if all sets of parameters are assumed to be identical across individuals and time. Analysis of covariance (ANCOVA) is used when the intercepts are assumed to be different but fixed, while error component modelling (ECM) is applicable for different random intercepts. The seemingly unrelated regression (SUR) technique is used to estimate different sets of fixed parameters, while estimating different sets of random parameters requires using random coefficient modelling (RCM).

2.3 The Growth Curve Literature

Individual monitoring is a common problem in biological research. Therefore, various models and techniques have been investigated to analyze growth curves. In the biological

literature, time-series data are usually called longitudinal data or repeated measurement data.

Pothoff and Roy (1964) suggested the multivariate analysis of variance (MANOVA) model to estimate growth curves. The MANOVA model requires that all individuals be measured at the same age. This model allowed confidence interval estimation and hypothesis testing. Rao (1965) extended Pothoff and Roy's model, transforming it to a random coefficient model where the matrix of independent variables were identical across individuals (*i.e.*, $\mathbf{X}_i = \mathbf{X}$ for any i). Grizzle and Allen (1969) used an analysis of covariance technique where the weighting was obtained from a subset of covariates. Geisser (1970) provided a Bayesian analysis of the MANOVA model.

Fearn (1975) applied the general Bayesian linear model proposed by Lindley and Smith (1972) to growth curves. He also studied predictions given a sample from this model. Rao (1975) gave an empirical Bayes solution to a random coefficient model where the parameters of the individual models were assumed randomly drawn from a common distribution whose expected value was the average of the individual parameters. Laird and Ware (1982), considering the individual parameters as missing data, used the EM algorithm (Dempster *et al.* 1977) to estimate them. With the EM algorithm, slow convergence or convergence to a local rather than global maximum are major concerns. Berkey (1982) extended Rao's (1975) technique to nonlinear models. Hui and Berger (1983) also used an empirical Bayesian approach to growth curves. Their technique could be called an empirical Bayes estimation of a random coefficient model. Rao (1987) discussed the problem of predicting future observations with growth models. He analyzed parametric and nonparametric models. For the parametric models, he considered both Bayesian and empirical Bayesian techniques to deal with unknown parameters.

Bayesian and empirical Bayesian techniques have predominated biological growth curve literature. Without starting a philosophical debate on how they differ and which

one, if any, is more appropriate, one can notice that both are partially based on some auxiliary information. This auxiliary information, particular to an individual, individualizes a general model and its prediction of future observations. The prediction of future observations is dependent on the outcome of this individual-specific auxiliary information and on an estimate of the general model.

2.4 Conditional Densities and Expectations

Many techniques for analyzing remeasured individuals and for predicting future observations described so far assume that future observations are conditional on past, random, observations. The modeling process, therefore, involves conditional densities and expectations. A conditional density is a probability density function (pdf) where the outcome of a random variable X depends on the outcome of another random variable Y . The variable of interest is thus X given Y , or $X | Y$, and its pdf is noted $p(X | Y)$. Most of the theory concerning conditional pdf's was developed by Bayes (1763), and a brief summary is presented here.

Some basic matrix inversion rules are useful for deriving conditional densities and expectations. These rules can be found in certain texts (see Smith 1973, for instance).

$$\text{Rule (1)} \quad (\mathbf{D} + \mathbf{EFE}')^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{E}(\mathbf{E}'\mathbf{D}^{-1}\mathbf{E} + \mathbf{F}^{-1})^{-1}\mathbf{E}'\mathbf{D}^{-1}$$

$$\text{Rule (2)} \quad (\mathbf{D} + \mathbf{B})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}(\mathbf{D}^{-1} + \mathbf{B}^{-1})^{-1}\mathbf{D}^{-1}$$

$$\text{Rule (3)} \quad (\mathbf{D} + \mathbf{B})^{-1}\mathbf{B} = \mathbf{I} - (\mathbf{D} + \mathbf{B})^{-1}\mathbf{D}$$

Conditional densities can be defined for univariate or multivariate variables. The multivariate case is more general. Let \mathbf{X} and \mathbf{Y} be two random vectors of n elements with pdf $p(\mathbf{X})$ and $p(\mathbf{Y})$, respectively. Their joint pdf is $p(\mathbf{X}, \mathbf{Y})$. Using Bayes' theorem on conditional probabilities, the joint pdf can be rewritten:

$$p(\mathbf{X}, \mathbf{Y}) = p(\mathbf{X} | \mathbf{Y})p(\mathbf{Y})$$

$$= p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})$$

From this theorem, the conditional pdf of \mathbf{X} given \mathbf{Y} can be defined as:

$$p(\mathbf{X}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{Y})}.$$

When \mathbf{Y} is given, the denominator $p(\mathbf{Y})$ contains no unknown. It can be seen as a proportionality constant, so that the multiple integral of $p(\mathbf{X}|\mathbf{Y})$ over X_1, \dots, X_n yields one. This constant will be equal to the inverse of the integral of the numerator over the domain of \mathbf{X} :

$$p(\mathbf{Y}) = 1 / \int_{X_1} \int_{X_2} \dots \int_{X_n} p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})dX_1 dX_2 \dots dX_n$$

For instance, if \mathbf{Y} given \mathbf{X} is distributed as a $N(\mathbf{X}, \mathbf{\Omega})$,

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{(2\pi)^{n/2} |\mathbf{\Omega}|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{Y} - \mathbf{X})'\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}) \right]$$

and \mathbf{X} is distributed as a $N(\mu, \mathbf{\Delta})$,

$$p(\mathbf{X}) = \frac{1}{(2\pi)^{n/2} |\mathbf{\Delta}|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{X} - \mu)'\mathbf{\Delta}^{-1}(\mathbf{X} - \mu) \right]$$

the unconditional variable \mathbf{Y} will be distributed as a $N(\mu, \mathbf{\Delta} + \mathbf{\Omega})$:

$$p(\mathbf{Y}) = \frac{1}{(2\pi)^{n/2} |\mathbf{\Delta} + \mathbf{\Omega}|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{Y} - \mu)'(\mathbf{\Delta} + \mathbf{\Omega})^{-1}(\mathbf{Y} - \mu) \right].$$

To show this, rather than using the integral of the product of $p(\mathbf{Y}|\mathbf{X})$ and $p(\mathbf{X})$, an easy alternative can be used. The variable \mathbf{Y} given \mathbf{X} can be seen as the sum of the variable \mathbf{X} , normally distributed, plus an error term, \mathbf{e} , uncorrelated with \mathbf{X} , normally distributed too, with expected value zero and variance $\mathbf{\Omega}$. Obviously, \mathbf{Y} will be normally distributed since the sum of normal variables yields a normal variable. The expected value and the variance of this normal variable can be found easily:

$$\mathbf{Y} = \mathbf{X} + \mathbf{e},$$

$$E[\mathbf{Y}] = E[\mathbf{X}] + E[\mathbf{e}] = \boldsymbol{\mu},$$

$$\text{var}(\mathbf{Y}) = \text{var}(\mathbf{X}) + \text{var}(\mathbf{e}) = \boldsymbol{\Delta} + \boldsymbol{\Omega}.$$

Using the definition of a conditional density, it can be shown that the variable \mathbf{X} given \mathbf{Y} will also be normally distributed:

$$\begin{aligned} p[\mathbf{X}|\mathbf{Y}] &= \frac{\frac{1}{(2\pi)^n |\boldsymbol{\Delta}|^{1/2} |\boldsymbol{\Omega}|^{1/2}} \exp \left[-\frac{1}{2} \{ (\mathbf{Y} - \mathbf{X})' \boldsymbol{\Omega}^{-1} (\mathbf{Y} - \mathbf{X}) + (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \} \right]}{\frac{1}{(2\pi)^n |\boldsymbol{\Delta} + \boldsymbol{\Omega}|^{1/2}} \exp \left[-\frac{1}{2} \{ (\mathbf{Y} - \boldsymbol{\mu})' (\boldsymbol{\Delta} + \boldsymbol{\Omega})^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \} \right]} \\ &= \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Delta}|^{1/2} |\boldsymbol{\Omega}|^{1/2}} \times \\ &\quad \exp \left[-\frac{1}{2} \{ (\mathbf{Y} - \mathbf{X})' \boldsymbol{\Omega}^{-1} (\mathbf{Y} - \mathbf{X}) + (\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Delta}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right. \\ &\quad \left. - (\mathbf{Y} - \boldsymbol{\mu})' (\boldsymbol{\Delta} + \boldsymbol{\Omega})^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \} \right] \\ &= \frac{1}{(2\pi)^{n/2} |(\boldsymbol{\Delta}^{-1} + \boldsymbol{\Omega}^{-1})^{-1}|^{1/2}} \times \\ &\quad \exp \left[-\frac{1}{2} \{ (\mathbf{X} - E[\mathbf{X}|\mathbf{Y}])' (\boldsymbol{\Delta}^{-1} + \boldsymbol{\Omega}^{-1}) (\mathbf{X} - E[\mathbf{X}|\mathbf{Y}]) \} \right] \end{aligned}$$

where

$$E[\mathbf{X}|\mathbf{Y}] = (\boldsymbol{\Delta}^{-1} + \boldsymbol{\Omega}^{-1})^{-1} \boldsymbol{\Omega}^{-1} \mathbf{Y} + (\boldsymbol{\Delta}^{-1} + \boldsymbol{\Omega}^{-1})^{-1} \boldsymbol{\Delta}^{-1} \boldsymbol{\mu}.$$

The expected value of the conditional pdf \mathbf{X} given \mathbf{Y} is a weighted average between the expected values of the variables \mathbf{Y} given \mathbf{X} and \mathbf{X} , where the weights are inversely proportional to their respective variance. The variance of \mathbf{X} given \mathbf{Y} is the inverse of the sum of the inverses of the respective variances. The expected value $E[\mathbf{X}|\mathbf{Y}]$ can be seen to be a random variable since it depends on the random outcome of \mathbf{Y} . It is the sum of a constant, $(\boldsymbol{\Delta}^{-1} + \boldsymbol{\Omega}^{-1})^{-1} \boldsymbol{\Delta}^{-1} \boldsymbol{\mu}$, and a random vector normally distributed, \mathbf{Y} , multiplied by a vector of constants. Therefore, it will be normally distributed, with the

following expected value and variance:

$$E[E[\mathbf{X} | \mathbf{Y}]] = E[\mathbf{X}] = \boldsymbol{\mu}$$

$$\text{var}(E[\mathbf{X} | \mathbf{Y}]) = \text{var}(\mathbf{X}) - E[\text{var}(\mathbf{X} | \mathbf{Y})] = \boldsymbol{\Delta} - (\boldsymbol{\Delta}^{-1} + \boldsymbol{\Omega}^{-1})^{-1}.$$

If it is assumed that $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$ are two positive definite matrices, then it can be shown that $\text{var}(\mathbf{X})$ is always larger than $\text{var}(E[\mathbf{X} | \mathbf{Y}])$. A matrix \mathbf{A} is larger than another matrix \mathbf{B} if the matrix $\mathbf{A} - \mathbf{B}$ is positive definite.

Using matrix inversion rules (2) and (3) defined above, the expected value of the conditional density is equivalent to

$$E[\mathbf{X} | \mathbf{Y}] = \boldsymbol{\Delta}(\boldsymbol{\Delta} + \boldsymbol{\Omega})^{-1}\mathbf{Y} + \boldsymbol{\Omega}(\boldsymbol{\Delta} + \boldsymbol{\Omega})^{-1}\boldsymbol{\mu}.$$

As demonstrated, statistical computations with conditional variables follow the same general rules as with unconditional variables, despite the more elaborate algebra. It is important to be familiar with the algebra of conditional variables to understand models making use of past observations to predict future observations.

2.5 Overview

Permanent sample plot attributes can be assumed to be explained by some general biological phenomenon. Observations of these attributes on a given plot at a point in time will also be influenced by the individual character of the plot and time-related variables, such as climate. Repeated measurements on a group of plots provide information on both the influence of the individual character and the time period effect. This suggests that permanent sample plot observations can be explained by a general model plus random terms specific to the individual and the time period.

Since future observations must be predicted on a given plot, a prediction of the individual character of the plot as well as an estimate of the general model are needed

to individualize the volume prediction. This approach recognizes the random nature of the relationship between observations coming from a sample plot. Future observations depend on the general model, and the outcome of the individual plot character. The individual plot character can be predicted from past observations. Therefore, predictions of future observations are dependent on past observations.

Many approaches can be used to predict future observations based on past observations. The approach suggested in this dissertation was inspired by the empirical Bayes models described by Rao (1975, 1987) and Hui and Berger (1983), the random coefficient model (Swamy 1970, 1971, 1974; Amemyia 1978; Ferguson and Leech 1978; Biging 1985), as well as the theory related to error component modeling.

Chapter 3

Statistical Methods

Future volumes can be expressed as the current volume plus a volume increment. The analogy with the interest on the money in a bank account is often used to explain forestry ideas such as volume growth and allowable annual cut to non-foresters. Going from interest to interest rate to biological growth rate are steps easy to understand, and the mathematics relevant to interest rates can simply be adapted to a forestry context. Consequently, the parameter of interest in the proposed model is the growth rate. A biologically-reasonable model for predicting volume growth rate is defined and estimated. This predicted volume growth rate is used to compound the current volume of a plot, the same way money can be compounded with a certain interest rate.

This chapter is divided into four parts. In the first section, the similarities between biological growth rate and interest rate are analyzed and explained. In the second section, a two-stage statistical model based on interest rate mathematics and fundamental principles of biological growth is suggested to predict volume growth rate. Parameter estimation is discussed and estimators are proposed in the third section. Finally, algorithms for compounding the current volume of a sample plot and for computing a confidence interval around the predicted volumes are presented in the last section.

3.1 The Timber Capital

The trees on a permanent sample plot can be compared to a timber capital destined for yielding interest. The mathematics relevant to interest rates are simple and well-known;

converting them to growth rates is straightforward as it is shown in section 3.1.1 The main difference between money and trees is the uncertainty coming with observations on biological organisms such as trees. This uncertainty must be recognized and incorporated into the mathematical expression of the compounded capital. In section 3.1.2, the problem of uncertainty attached to volume observations is addressed.

3.1.1 The Interest Rate Analogy

The growth in volume for a tree in a particular year is the amount of volume the tree gained between the beginning and the end of the growing season. When the growth is divided by the initial volume, it can be seen as a growth rate (*i.e.*, the amount of volume per unit of initial volume gained in a year). The growth rate is the annual “interest rate” on the initial volume of a particular tree in a given year. This image can be extended to a plot or a stand without problem. In this section, the mathematics relevant to interest rates and their similarities with growth rates are explained.

The interest rate at a particular bank will likely fluctuate over time. It will also likely differ between banks; two different banks can have different interest rates. In the same way, the growth rate of a tree (or a plot or stand) will be specific to a given individual and time. Mathematically, this gives

$$V_{i(j+1)} = V_{ij}(1 + \rho_{ij}) \quad (3.1)$$

where

V_{ij} = volume measured on individual i at time j ,

ρ_{ij} = annual growth rate on individual i at time j , and

$(1 + \rho_{ij})$ = annual multiplicative rate.

In forestry, permanent sample plots are commonly used to monitor growth. However, annual measurements on permanent sample plots are rarely available. The sample plots

are remeasured at a regular frequency, usually between three and ten years. When the time interval between measurements j and $j + 1$ is more than one growing season, periodic growth is then observed. In this case, the ratio between the volume at the end of the period and the initial volume is the periodic multiplicative rate. The periodic multiplicative rate is the product of the annual multiplicative rates. Since the annual multiplicative rates are not observed, the periodic multiplicative rate can be considered as the geometric mean of the annual multiplicative rates raised to a power, L_j , equal to the time lag between measurements j and $j + 1$:

$$V_{i(j+1)} = V_{ij}(1 + \tilde{\rho}_{ij})^{L_j} \quad (3.2)$$

where $\tilde{\rho}_{ij} = [\Pi_{(j=1)}^{L_j}(1 + \rho_{ij})]^{\frac{1}{L_j}} - 1$.

The periodic multiplicative rate is the percentage of initial volume obtained after a certain period of time. Therefore, the notation used for annual remeasurements is somewhat inadequate for periodic remeasurements since the subscript j in $\tilde{\rho}_{ij}$ refers to the period between measurements j and $j + 1$ while in V_{ij} it refers to the instantaneous moment at measurement j . Model (3.2) can be rewritten using the subscript p to refer to the time interval between measurements j and $j + 1$ with V_{ip}^0 the initial volume on plot i and V_{ip}^1 the volume at the end of the period:

$$V_{ip}^1 = V_{ip}^0(1 + \tilde{\rho}_{ip})^{L_p}. \quad (3.3)$$

With this notation, the growth rate is located in space and time. The subscript i indicates a geographic space, (delimited by some longitude and latitude) and the subscript p is temporal space (delimited by an initial year and a final year).

This approach to growth modeling could be taken for other stand attributes, such as diameter, height, basal area or number of stems per hectare. For illustrative purposes, the approach is shown for plot volume. For the remainder of this thesis, the term volume will refer to plot volume.

The periodic multiplicative rate in model (3.3) can also be expressed using the Napierian base ($e = 2.7182818\dots$)

$$V_{ip}^1 = V_{ip}^0 \times \exp[\theta_{ip}L_p] \quad (3.4)$$

where $\theta_{ip} = \ln(1 + \tilde{\rho}_{ip})$.

The parameter θ_{ip} is easily isolated

$$\theta_{ip} = \frac{1}{L_p} \ln \left[\frac{V_{ip}^1}{V_{ip}^0} \right] \quad (3.5)$$

Using the logarithm rules and the definition of a derivative, it can be shown that, as the time period goes to zero, θ_{ip} becomes the instantaneous growth rate

$$\begin{aligned} \lim_{L_p \rightarrow 0} \theta_{ip} &= \lim_{L_p \rightarrow 0} \frac{1}{L_p} \ln \left[\frac{V_{ip}^1}{V_{ip}^0} \right] \\ &= \lim_{L_p \rightarrow 0} \frac{\ln[V_{ip}^1] - \ln[V_{ip}^0]}{L_p} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Delta \ln[V]}{\Delta t} \\ &= \frac{\partial \ln[V]}{\partial t} \\ &= \frac{1}{V} \frac{\partial V}{\partial t} \end{aligned} \quad (3.6)$$

where t is time.

Similarly, $\theta_{ip}L_p$ can be shown to be the integral of θ_{ip} with respect to time between the initial and the final measurements. Time can be expressed on a relative scale. The initial measurement is taken at time 0, and the final one is taken at time L_p . This yields

$$\theta_{ip}L_p = \int_0^{L_p} \theta_{ip} dt \quad (3.7)$$

which is the periodic multiplicative rate expressed in a logarithmic scale.

The concept of growth as the derivative of volume with respect to time (or volume as the integral of growth) was recognized in the early sixties (Buckman 1962, Clutter

1963). This is a simplification of reality; it assumes that growth is a constant process, which is not exactly true since growth stops in winter. A discrete analog would be to take the annual interest rate analogy. This approach has some shortcomings too; there is no growth assumed before the end of the growing season, similar to the annual interest on a bank account that is paid once a year. Another discrete approach would be to follow the modern bank system and use daily interest rate, but this would demand a tremendous data set. Assuming that growth is a continuous process leads to relatively easy mathematical computations while being close to reality.

3.1.2 Uncertainty and Growth Rate

The deterministic model (3.4) could be used if the true initial and final volumes were observed. However, these quantities will likely be measured with some error. The observed volumes are random variables. The variability in observed volumes will introduce some variation into the observed growth rate. Because the growth rate is essentially the ratio of two variables, it becomes a random variable itself. A stochastic term must be incorporated in model (3.4) to account for this variability. *A priori* assumptions are then made about this error term in order to draw inferences using statistical theory after mathematically fitting the model. This stochastic term can be included in different ways. Three common ones are

$$(V_{ip}^1/V_{ip}^0) = \exp[\theta_{ip}L_p] + u_{ip} \quad (3.8)$$

$$(V_{ip}^1/V_{ip}^0) = \exp[(\theta_{ip}L_p) + u_{ip}] \quad (3.9)$$

$$(V_{ip}^1/V_{ip}^0) = \exp[(\theta_{ip} + u_{ip})L_p]. \quad (3.10)$$

The choice of a model depends on its mathematical tractability and its biological interpretation. Model (3.8) is a nonlinear model and is particularly difficult to fit to recent permanent sample plots. Model (3.9) is linear after a logarithmic transformation.

It adds an error term to the periodic growth rate. The error term will undoubtedly depend on the length of the time period even if not explicitly expressed in the model. Model (3.10) is also log-linear. The error term is included with the instantaneous growth rate. Its dependency on the time lag is therefore implicitly expressed in the model. This model has the advantages of being simple to fit and of corresponding to the interpretation of a growth rate whose variation is dependent on the period length. For these reasons, model (3.10) was considered the best approach.

Bank account and trees have this point in common: they become bigger as time passes. However, while knowing the exact amount of money in a bank account is no problem, direct observations of plot volume is impossible. An error term can be included in the mathematical expression of periodic growth rate to account for this uncertainty. Model (3.10) is the basic expression of growth rate with uncertainty. Within this basic expression, it is possible to assume that the growth rate, θ_{ip} , can be explained by some biological phenomena. A relationship explaining θ_{ip} from some variables can be substituted in the basic expression. This framework leads to the statistical model to predict future growth rates, which is discussed in the next section.

3.2 The Statistical Model

The framework proposed for estimating the statistical model is a two-stage procedure. The first stage is the basic expression describing the mathematics of growth rate, and the second stage is a biological explanation of the true growth rate.

The framework was designed specifically for the irregular remeasurements of permanent sample plots. These irregular remeasurements, along with overlapping time periods will create notation difficulty. Conventions must be defined to avoid confused notation. Once the notation is defined, each stage of the procedure can be discussed. The first

stage deals with the problem of a small number of observations on each plot and the error components. The second stage makes use of Von Bertalanffy's growth curve to estimate the growth rate from site index and volume. Finally, both stages are joined in the combined model.

3.2.1 Notational Conventions

It will be assumed that there are N plots covering M intervals. Some time intervals may be overlapping. For instance, if some plots were measured in 1970, 1975 and 1980, and the other plots were measured in 1973, 1978, and 1983, there would be four time intervals: (1970, 1975], (1973, 1978], (1975, 1980], and (1978, 1983]¹. Not every plot will have an observation in each time interval. Furthermore, the number of plots measured each year will vary with the money budgeted annually for the remeasurement of permanent sample plots. Consequently, the number of remeasurements will differ among plots and among time periods. It will be assumed that there are T_i growth observations on plot i ($i = 1, \dots, N$), S_p observations in time interval p ($p = 1, \dots, M$), and W observations in total where

$$W = \sum_{i=1}^N T_i = \sum_{p=1}^M S_p; \quad \text{with} \quad W > N \quad \text{and} \quad S_p > 1.$$

Because time intervals may be overlapping, consecutive measurements on a plot do not lead to observations in consecutive time intervals. For the example mentioned above, the first set of plots would have observations in time intervals (1970, 1975] and (1975, 1980], while the second set of plots would have observations in time intervals (1973, 1978] and (1978, 1983]. The time interval numbering and the plot numbering are arbitrary. They are only coordinates to link an observation with geographic and time locations. Also, it is assumed that all plots used in the analysis come from the same geographic region, and

¹This notation supposes that the plots were measured at the end of the growing season.

are dominated by the same species. No notation is used to identify these two qualitative variables.

3.2.2 First-Stage Model

The first-stage model is the basic expression for periodic growth rate, with uncertainty included with the instantaneous growth rate. Recent permanent sample plots only have a few growth observations. Using an approximation, the proposed first-stage model can be estimated from plots that have had only two growth observations. This approximation is explained in the model definition section. Next, the error term is decomposed to take advantage of the measurements of many individuals in a given time interval. Assumptions about the error term and their justifications follow. Finally, the matrix notation of the model is given to simplify the notation and derivation of the parameter estimators.

3.2.2.1 Model Definition

The observed periodic multiplicative rate is the ratio V_{ip}^1/V_{ip}^0 . The average annual multiplicative rate observed during this time period is the observed periodic multiplicative rate raised to the power $1/L_p$. Because the model (3.10) is in exponential form, it will be more convenient to work with the natural logarithm of this ratio. So let

$$\begin{aligned} Y_{ip} &= \ln \left[\left(\frac{V_{ip}^1}{V_{ip}^0} \right)^{\frac{1}{L_p}} \right] \\ &= \frac{1}{L_p} \ln \left[\frac{V_{ip}^1}{V_{ip}^0} \right]. \end{aligned} \quad (3.11)$$

It will be assumed that the variable Y_{ip} represents an estimate of the instantaneous growth rate, θ_{ip} , since L_p is usually small. It is the sum of the true instantaneous growth rate plus an error term:

$$Y_{ip} = \theta_{ip} + u_{ip} \quad (3.12)$$

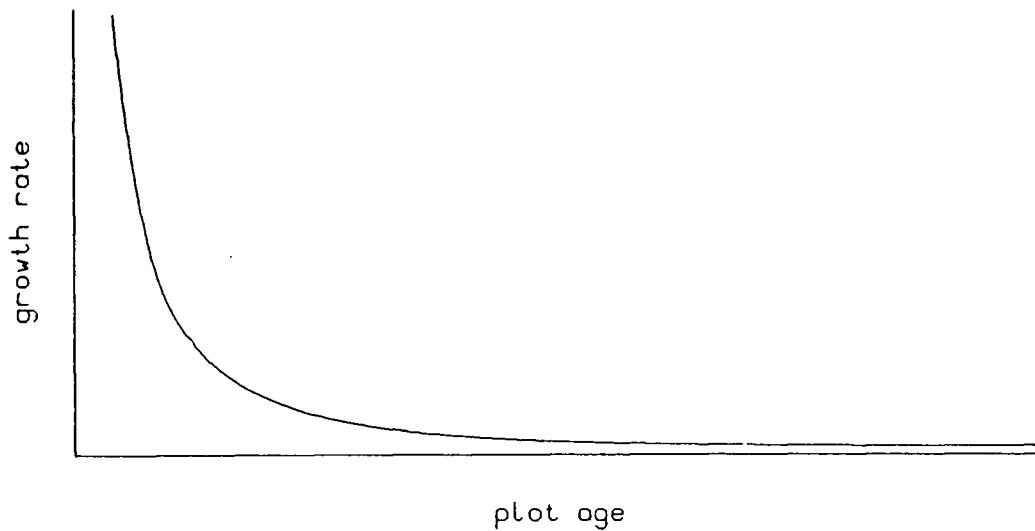


Figure 3.1: Growth Rate over Time for a Sigmoidal Yield Curve

which corresponds to the logarithmic transformation of model (3.10).

Recent permanent sample plots cover only a few time intervals, and if these time intervals are short, the whole period covered by the sample plot can still be considered close to zero, especially when it is compared to the lifespan of the stand. Therefore, the instantaneous growth rate could also be estimated from the ratio between the last and the first measurements on a sample plot. This corresponds to assuming a constant growth rate on a plot. Hence, the growth rates estimated in the intervals can be considered as repeated observations of the same parameter, the constant growth rate.

If the yield is assumed to follow a sigmoidal shape over time, the growth rate in the early ages declines rapidly to level off as the stand gets older (Figure 3.1). Therefore, the assumption of constant growth rate seems reasonable for a large part of the plot's life. Obviously, this assumption is a compromise to approximate the trend of growth rate over time for a sigmoidal yield curve and to allow for parameter estimation.

This approximation will simplify the model, but also bring some limitations. It likely could cause some bias. Assuming a constant growth rate on a plot will tend to underestimate the growth rate at the first measurement and overestimate it at the last measurement. This bias should be negligible, or at least reasonably small, if the whole period covered by the plot is short enough. The period of time where the approximation is reasonable depends on the species and the age of the plot. The bias should be more pronounced on young plots. Also, the rate of change in the instantaneous growth rate is species-dependent; fast growing species will level off faster than slow growing species. Considering these limitations, the approximation can still prove useful in many cases such as older plots and fast growing species. Model (3.12) can be approximated by

$$Y_{ip} = \theta_i + u_{ip} \quad (3.13)$$

This model can be estimated with as few as two growth observations on a permanent sample plot. That is, the model requires a minimum of three measurements on a permanent sample plot since the number of measurements is equal to the number of growth observations plus one. Also, measurements on different individuals in the same time period will allow decomposing u_{ip} , the error term.

3.2.2.2 Error Components

Following the principles of error component modeling, when measured on a given individual (*i.e.*, the individual is fixed), the error term u_{ip} can be divided into two components: the error due to the time-dependent variables unaccounted for in the model and a term for the unaccounted variables that are not time-dependent. This gives

$$u_{ip} = \tau_p + v_{ip} \quad (3.14)$$

where

r_p = error term explained by the time-dependent variables unaccounted for in the model.

It remains constant for all individuals in a time period but varies among time periods.

v_{ip} = unexplained residual noise. It accounts for the observed difference between individuals with the same true growth rate in a given time period.

Some unobservable or unmeasurable variables will have a similar effect on geographically close plots in a time period (Gregoire 1987). For example, climatic variables, and major insect and disease epidemics will uniformly affect a small region in a given time period. Total number of days in the growing season, rainfall, quantity and quality of sunlight, *etc.* should be roughly similar for all plots in a time period. However, it would be impossible to include all these variables in the model. Their cumulative effect will vary from time period to time period. With a two-way lay-out such as a panel data set, it becomes possible to estimate the variation of the random effect caused by these variables. It is therefore possible to distinguish the time-dependent ignorance from the unexplained ignorance. The error term in model (3.13) was decomposed to represent this knowledge about the stochastic variation in the model.

3.2.2.3 Assumptions

Combining equations (3.13) and (3.14), the model for estimating the growth rate on plot i becomes:

$$Y_{ip} = \theta_i + r_p + v_{ip}. \quad (3.15)$$

Assumption 1. $E[r_p] = E[v_{ip}] = 0$

Assumption 2. $E[r_p r_q] = \delta^2$ if $p = q$
 $= \delta_{pq}$ if $p \neq q$

$$\begin{aligned} \text{Assumption 3. } E[v_{ip}v_{jq}] &= \sigma_p^2 \text{ if } i = j \text{ and } p = q \\ &= 0 \text{ otherwise} \end{aligned}$$

$$\text{Assumption 4. } E[r_p v_{iq}] = 0$$

Assumption 5. r_p and v_{ip} are normally distributed.

These assumptions imply that u_{ip} is normally distributed and

$$\begin{aligned} E[u_{ip}] &= 0 \\ E[u_{ip}u_{jq}] &= \sigma_p^2 + \delta^2 \text{ if } i = j \text{ and } p = q \\ &= \delta^2 \text{ if } i \neq j \text{ and } p = q \\ &= \delta_{pq} \text{ if } p \neq q \end{aligned}$$

Assumption 1 means that the expected value of the observed growth rate on a plot is assumed to be the true growth rate for this plot. Since Y_{ip} is the logarithm of a ratio of random variables raised to a certain power, this assumption will make sense if and only if the expected value of the logarithm of the plot volume measured on plot i at time j is the logarithm of the true volume

$$E[Y_{ip}] = \theta_i \iff E[\ln V_{ij}] = \ln v_{ij}$$

where v_{ij} is the true volume. The error of the logarithm of the observed volume has often been assumed to have zero mean in growth and yield literature (Schumacher 1939, Clutter 1963, Curtis 1967, Ferguson and Leech 1978, Gregoire 1987), and the same assumption is made here. The bias created by the approximation of a constant growth rate is assumed to be negligible.

Assumption 2 concerns the time-dependent error terms affecting the growth rate. The effects of these error terms should be correlated with each other since the time

periods are overlapping. Following Garcia (1983), it can be assumed that the correlation is proportional to the overlap between two time periods. Nonoverlapping time periods are considered independent; the weather varies randomly from one period to another. Overlapping time periods share the same climatic conditions for a certain number of years which should indicate the presence of correlation. This assumption also implies that consecutive growth observations on a given individual are considered independent because they are always taken in nonoverlapping time periods.

Let C_{pq} be the number of overlapping years between periods p and q , then the covariance between r_p and r_q , δ_{pq} , can be assumed to be

$$\delta_{pq} = \delta^2 \frac{C_{pq}(L_p + L_q)}{2L_p L_q} \quad (3.16)$$

where L_p and L_q are the respective lengths of periods p and q .

To explain this estimator, let r_p be divided into two parts: r_{p1} and r_{p2} where r_{p1} is the temporal error term for the first $L_p - C_{pq}$ years, and r_{p2} is the temporal error term for the remaining C_{pq} years. Similarly, let r_q be divided into r_{q1} and r_{q2} where r_{q1} is the temporal error term for the first C_{pq} years, and r_{q2} is the temporal error term for the remaining $L_q - C_{pq}$ years. Both r_{p2} and r_{q1} are error terms for the same time period. Within a time period, the variance is assumed constant. This means that the variance of r_{p2} is $C_{pq}\delta^2/L_p$ when estimated from time interval p . Similarly, the variance of the error explained by the same time interval but estimated from time interval q , r_{q1} , would be $C_{pq}\delta^2/L_q$. Also, the correlation between r_{p2} and r_{q1} is really the variance of the time-dependent error term for the C_{pq} overlapping years. Therefore, $C_{pq}\delta^2/L_p$ and $C_{pq}\delta^2/L_q$ are two different estimates of the correlation between r_{p2} and r_{q1} . Averaging out both estimates will yield a single estimate of the correlation between the two error terms, providing in the same time an estimate of the variance of the time-dependent error term for these C_{pq} overlapping years.

Mathematically, this yields

$$\begin{aligned}
 E[r_p r_q] &= E[(r_{p1} + r_{p2})(r_{q1} + r_{q2})] \\
 &= E[r_{p1} r_{q1}] + E[r_{p1} r_{q2}] + E[r_{p2} r_{q1}] + E[r_{p2} r_{q2}] \\
 &= E[r_{p2} r_{q1}] \\
 &= \frac{1}{2}(E[r_{p2}^2] + E[r_{q1}^2]) \\
 &= \frac{1}{2}(\text{var}(r_{p2}) + \text{var}(r_{q1})) \\
 &= \frac{C_{pq}}{2} \left[\frac{\delta^2}{L_p} + \frac{\delta^2}{L_q} \right] \\
 &= \delta^2 \frac{C_{pq}(L_p + L_q)}{2L_p L_q} \tag{3.17}
 \end{aligned}$$

Two nonoverlapping time periods have $C_{pq} = 0$ satisfying the assumption of independence; two completely overlapping time periods have $L_p = L_q = C_{pq}$ which would yield a covariance of δ^2 . Assuming a known form for δ_{pq} is convenient since it decreases the number of unknown parameters to be estimated in the model.

Assumption 3 implies that the unexplained error terms are independently distributed with common variance within a time period. This error term represents the stochastic variation between individuals observed in given time period after correction for the true growth rate. It accounts for the general ignorance in the model.

For various reasons, some plots can be more sensitive than others in unusual but not extreme climatic conditions. For example, rainfalls slightly below the average can have serious effects on some plots and leave other plots almost unaffected. Other special conditions can prevail on a particular stand in a time period. A local insect infestation, for example, could affect only a small number of plots. For extreme conditions, it can effect all plots similarly. For instance, a major insect infestation would probably have the same kind of impact on all plots. The variance of this error term is assumed to differ

among time periods. The variation among plots within a time period will differ with the magnitude of the effect of climatic conditions on each plot. Some climatic conditions will cause a wide range of impact on individual plot growth rates while other conditions will affect all plots in a similar fashion.

This error term can be seen as the interaction between time periods and plots, and it is assumed that the variance of this interaction varies with time periods. It is also assumed that the interaction between a given plot and a time period is uncorrelated with the interaction between the same plot and another time period, or with the interaction between another plot and the same time period.

Assumption 4 implies that the time-dependent error term is uncorrelated with the unexplained error term. The individual plot behaviour in certain climatic conditions is unrelated with the average behaviour of all plots in the same conditions.

Assumption 5 was justified by Aitchison and Brown (1957; p. 1-2) who mentioned that a variable which is the product of numerous independent random variables tends to be lognormally distributed around its true value. If the growth rate at any time is considered a random proportion of the initial volume, and if the growth rates are considered independent for a given individual, then the volume tends to be lognormally distributed. Therefore, the observed growth rate tends to be normally distributed.

3.2.2.4 Matrix Notation

Model (3.15) and its assumptions can be expressed with matrices. To define the model using matrix notation, the observations are sorted initially by time periods and secondly by plots. The first observation is then the observation measured in period 1 on the plot with the smallest plot number from all plots measured in period 1, and is denoted Y_{s1} .

The last observation is the observation measured in period M on the plot with the largest plot number from all plots measured in period M (Y_{lM}). There are W observations in the matrix containing the growth observations.

The first-stage model is conditional on θ . The second-stage model will express θ_i as a function of site index and initial plot volume. This approach will make the estimation of the covariance matrix of the random terms easier. The first-stage model is:

$$\mathbf{Y} = \mathbf{F}\theta + \mathbf{G}\mathbf{r} + \mathbf{v} \quad (3.18)$$

with

$$\mathbf{Y} = \begin{bmatrix} Y_{s1} \\ \vdots \\ Y_{lM} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} r_1 \\ \vdots \\ r_M \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} v_{s1} \\ \vdots \\ v_{lM} \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & \dots & \mathbf{F}_N \end{bmatrix} \quad \mathbf{F}_i = \begin{bmatrix} F_{i1} \\ \vdots \\ F_{iW} \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \dots & \mathbf{G}_M \end{bmatrix} \quad \mathbf{G}_p = \begin{bmatrix} G_{p1} \\ \vdots \\ G_{pW} \end{bmatrix}$$

where $F_{ik} = 1$ if the k^{th} element of \mathbf{Y} was measured on plot i
 $= 0$ otherwise

and $G_{pk} = 1$ if the k^{th} element of \mathbf{Y} was measured in time period p
 $= 0$ otherwise.

The matrix of the sum of both error components would be $\mathbf{u} = \mathbf{G}\mathbf{r} + \mathbf{v}$. The various assumptions about the error terms lead to the following matrices

$$E[\mathbf{v}\mathbf{v}'] = \mathbf{\Sigma} = \text{diag} \left[\sigma_1^2 \mathbf{I}_{S_1} \quad \dots \quad \sigma_M^2 \mathbf{I}_{S_M} \right] \quad (3.19)$$

$$E[\mathbf{r}\mathbf{r}'] = \delta^2 \mathbf{\Lambda} \quad (3.20)$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & \lambda_{12} & \dots & \lambda_{1M} \\ \lambda_{21} & 1 & \dots & \lambda_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{M1} & \lambda_{M2} & \dots & 1 \end{bmatrix}$$

and

$$\lambda_{pq} = \frac{C_{pq}(L_p + L_q)}{2L_p L_q}$$

$$E[\mathbf{u}\mathbf{u}'] = \delta^2 \mathbf{G}\mathbf{\Lambda}\mathbf{G}' + \mathbf{\Sigma} = \mathbf{\Omega}. \quad (3.21)$$

This concludes the first-stage model. In this model, a constant growth rate is assumed for each plot. Since the individual is fixed, the first-stage model yields an estimate of the approximated constant growth rate of the plot. This estimate comes from the individual growth observations and can be called the observed instantaneous growth rate. Using a group of individuals, the growth rate can also be predicted from independent variables. This is the second-stage model.

3.2.3 Second-Stage Model

The second-stage model provides a biological relationship between the true instantaneous growth rate and the site index and the volume. This relationship is derived from Von Bertalanffy's growth curve.

The background explaining the relationship is presented first. The statistical model and its assumptions follow. Finally, the matrix notation of the model is given.

3.2.3.1 Background

It is reasonable to assume that the individual growth rate can be explained by some variables plus an individual-specific random term. For example, it can be assumed that the individual growth rate is a function of site index (SI) and initial volume (V). Any other factor that could have an influence on the individual growth rate can be included in a stochastic term. This function can be assumed to be identical for all individual growth rates, with fixed parameters and an additive error term

$$\theta_i = f(SI_i, V_i) + e_i. \quad (3.22)$$

Fitting the relationship expressed in (3.22) would provide an estimate of the growth rate of a plot as explained by its site index and its initial volume. Also, the stochastic nature of the individual growth rate is recognized by the additive error term.

The structure of the deterministic part of equation (3.22) can be derived from the biological interpretation leading to Von Bertalanffy's growth curve. The fundamental assumption underlying this curve is as follows. The potential growth is the difference between anabolic gains and catabolic losses. Anabolic gains are assumed to be proportional to the absorptive surface. The absorptive surface is assumed to be proportional to the volume raised to the power $2/3$. Catabolic losses are assumed to be proportional to the volume. This gives

$$\text{Potential growth} = \kappa_1 V^{2/3} - \kappa_2 V. \quad (3.23)$$

The potential growth is the production of surplus metabolic products (Pienaar and Turnbull 1973), which is the gain in biomass. The gain in biomass is distributed among various components of the trees: fine roots, large roots, stem, branches, leaves and bark. Therefore, the actual volume growth depends on the percentage of the total gain in biomass going to stem biomass. This percentage, κ_3 , is an efficiency coefficient relating volume

production to production of surplus metabolic products:

$$\begin{aligned}\text{Actual growth} &= \kappa_3(\kappa_1 V^{2/3} - \kappa_2 V) \\ \frac{\partial V}{\partial t} &= \kappa_3 \kappa_1 V^{2/3} - \kappa_3 \kappa_2 V.\end{aligned}\tag{3.24}$$

Keyes and Grier (1981) showed that the efficiency coefficient was correlated with site quality. Richer sites have a better availability of nutrients to put on stem growth. Poorer sites are poorer in nutrients, and more energy must be dedicated to fine roots production in order to get the nutrients. Kurz and Kimmins (1987) similarly found that the percentage of total biomass going into below-ground production is inversely proportional to site quality. As site quality decreases, a larger proportion of total biomass goes below ground, and consequently, a smaller proportion goes above ground. If the percentage of above-ground biomass is proportional to site quality, it can be expected that the percentage of biomass going into stem biomass will be correlated to site quality as well. Axelsson and Axelsson (1986) showed that the efficiency coefficient increases with fertilization. Richer sites do not need as developed of fine root system as the poorer sites, and are therefore more efficient at transforming their production of surplus metabolic products into growth. Site index can be assumed to be highly correlated with site quality; therefore, it is reasonable to assume that the efficiency coefficient is proportional to site index. Let κ_3 be $\kappa_4 SI$. Equation (3.24) can be rewritten

$$\begin{aligned}\frac{\partial V}{\partial t} &= \kappa_4 SI \kappa_1 V^{2/3} - \kappa_4 SI \kappa_2 V \\ &= \beta_1 SI V^{2/3} - \beta_2 SI V.\end{aligned}\tag{3.25}$$

The growth rate is the growth divided by the volume. Dividing both sides of equation (3.25) by V yields

$$\frac{1}{V} \frac{\partial V}{\partial t} = \beta_1 \frac{SI}{V^{1/3}} - \beta_2 SI.\tag{3.26}$$

3.2.3.2 Model Definition

The left-hand side of model (3.26) is the instantaneous growth rate, which is approximated by θ_i . Consequently, model (3.26) represents the deterministic part of model (3.22). The complete model (3.22) can be rewritten:

$$\theta_i = \beta_1 \frac{SI_i}{(V_i^0)^{1/3}} - \beta_2 SI_i + e_i. \quad (3.27)$$

The parameter θ_i is assumed to be constant during the time periods covered by plot i . However, this does not correspond to the concept of instantaneous growth rate. This concept can be retrieved in the second-stage model if θ_i is assumed to be measured at a single point, within the interval between the first measurement and the last measurement on plot i . The volume at this point can be used as the initial volume. This definition of initial volume will allow for simulating an instantaneous observation of the growth rate. Hui and Berger (1983) adopted a comparable strategy in a similar situation, mentioning that it should bring some robustness to the model.

The constant growth rate is an average of the instantaneous (but unknown) growth rates. Therefore, its value is the value of an instantaneous growth rate located between the first and the last measurements. This suggests that an “average” initial volume would lead to a more appropriate definition of initial volume for an instantaneous growth rate. There are many possible definitions of this “average” initial volume, and since the real growth rate is unknown, there seems to be no definite rule to define this initial volume. One possible estimator is to take the mean volume, that is the arithmetic average of the observed volumes on a plot.

$$V_i^0 \equiv \frac{\sum_j V_{ij}}{T_i + 1} \quad (3.28)$$

where V_{ij} is the volume observed on plot i at measurement j . When the growth rate is assumed to be estimated at this point, it better corresponds to the concept of instantaneous growth rate represented by θ_i .

3.2.3.3 Assumptions

Once an appropriate definition of the initial volume is accepted, stating the assumptions about the error terms is the next step. These assumptions are:

Assumption 6. SI_i and $(V_i^0)^{1/3}$ are fixed and measured without error

Assumption 7. $E[e_i] = 0$

Assumption 8. $E[e_i e_j] = \phi^2 / (V_i^0)^{1/3}$ if $i = j$
 $= 0$ otherwise

Assumption 9. e_i is normally distributed.

Assumption 6 is a common assumption in regression. As a consequence, the error term e_i is assumed to be uncorrelated with both independent variables, SI_i and V_i^0 .

Assumption 7 implies that the expected growth rate on a plot is assumed to be a function of site index and initial volume.

Assumption 8 has two implications. First, it implies that the growth rates are independent between plots. Knowing the true growth rate on a plot gives no idea about the true growth rate on another plot. The eighth assumption also states that the variance of the individual growth rates is inversely proportional to initial volume raised to a power of $1/3$. The variance of the individual growth rate should decrease as the plot gets older. Variation in growth rate is more important when the growth rate is large (*i.e.*, when the plots are young). If it is assumed that the variance is proportional to the expected individual growth rate

$$\text{var}(\theta_i) \propto \theta_i$$

(i.e., if the variance expressed as a proportion of the expected growth rate is assumed to be constant), it will also be inversely proportional to the initial volume raised to a power of $1/3$ since the expected value of the growth rate is itself inversely proportional to $1/(V_i^0)^{1/3}$.

Assumption 9 can be justified by invoking the Central Limit Theorem. The numerous variables influencing the growth rate unaccounted for in the second-stage model are assumed to be additive with their average effect believed to be negligible.

3.2.3.4 Matrix Notation

Model (3.27) and its assumptions can be written using matrix notation as:

$$\theta = \mathbf{X}\beta + \mathbf{e} \quad (3.29)$$

with the following matrices:

$$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} X_{11} & X_{21} \\ \vdots & \vdots \\ X_{1N} & X_{2N} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

where $X_{1i} = SI_i/(V_i^0)^{1/3}$ and $X_{2i} = -SI_i$. From the assumptions about the error term, the distribution of θ is easily found:

$$\theta \sim N(\mathbf{X}\beta, \phi^2 \mathbf{V}) \quad (3.30)$$

where \mathbf{V} is a diagonal matrix whose i^{th} diagonal element is $1/(V_i^0)^{1/3}$.

This concludes the second-stage model. This stage provides a biologically-reasonable model to predict the growth rate. The estimate given by the second-stage model can be seen as the predicted growth rate of a given plot.

3.2.4 Combined Model

The second-stage model (3.27) can be substituted into the first-stage model (3.15).

$$Y_{ip} = \beta_1 \frac{SI_i}{(V_i^0)^{1/3}} - \beta_2 SI_i + e_i + r_p + v_{ip} \quad (3.31)$$

3.2.4.1 Assumption

A new assumption is needed.

$$\text{Assumption 10. } E[v_{ip} | e_i] = E[r_p | e_i] = 0$$

Assumption 10 implies that the error terms from the first-stage model are uncorrelated with the error terms from the second-stage model. This means that the individual-dependent variables unaccounted for in the model are assumed to be uncorrelated with the unaccounted time-dependent variables or the residual error term. It should be noted that the error terms from the first-stage model are also uncorrelated with the independent variables since the independent variables are assumed to be fixed.

3.2.4.2 Matrix Notation

The combined model can be written with matrix notation as:

$$\mathbf{Y} = \mathbf{FX}\beta + \mathbf{Fe} + \mathbf{Gr} + \mathbf{v} \quad (3.32)$$

Based on the assumptions about the error terms, the distribution of \mathbf{Y} is:

$$\mathbf{Y} \sim N(\mathbf{FX}\beta, \Psi). \quad (3.33)$$

where $\Psi = \phi^2 \mathbf{FV}\mathbf{F}' + \Omega$.

3.2.5 Summary

Growth rate and interest rate are very similar in nature, and the mathematics relevant to interest rates can be adapted to permanent sample plot growth rates. On recent permanent sample plots, the growth rate can be approximated by a constant. When the time period effect is removed, an estimate of the constant growth rate can be obtained. This can be called the observed instantaneous growth rate. The instantaneous growth rate can also be predicted from independent variables using a relationship derived from Von Bertalanffy's growth curve. This relationship presents the growth rate as a function of site index and volume at the beginning of the growing season. The second-stage estimate of growth rate can be called the predicted instantaneous growth rate. Both models can be joined to express the growth rate observations as a function of initial volume, site index, an individual effect, a time effect and an unexplained error term. For predictions of future volumes on a permanent sample plot, the parameter of interest is the random variable θ_i , and its prediction is discussed in the next section.

3.3 Parameter Estimation

Traditionally, the parameter of interest has been β . However, this parameter does not recognize the individual character of each plot, represented by e_i . The parameter θ_i does recognize this individual character. The group of permanent sample plots provides the information to estimate the global parameter β . Repeated observations on a given plot yield the information to predict the individual term e_i of this plot. The estimate of β and the predictor e_i can be joined to predict θ_i , the actual parameter of interest.

It is a common technique in parameter estimation to assume that the covariance matrix of the model random terms are known in order to derive the parameter estimators.

This simplification often allows a better understanding of the estimators. The estimators of the parameters composing θ are first derived for the case where the covariance matrix of the model random terms are assumed to be known. Second, estimators for the unknown covariance matrices are defined. Finally, the estimated covariance matrices are substituted in the parameter estimators to yield estimates of the best linear unbiased predictor of θ and of its variance.

3.3.1 Estimation When Covariance Matrices Are Known

When the covariance matrix of the various random terms in the combined model (3.32) are assumed known, only the parameters β and e need to be estimated. These parameter estimates are random variables, and their variances can also be estimated. The parameter estimates are discussed first followed by a discussion of the estimate of the variance of the predictor of θ_i .

Deriving and simplifying the estimators involved a lot of matrix algebra. Most of the algebraic steps were carried out to provide a proof of the estimators. While complex at first glance, these steps are easy to follow using the matrix inversion rules given in section 2.4. Of course, the first and the final lines of any proof are sufficient to understand the text. The intermediate lines are complimentary, for the pleasure of the mathematically-inclined reader.

3.3.1.1 Parameter Estimators

Estimators for β and e are first derived from a theorem given by Harville (1976), and combined to form a predictor of θ . The simplification and interpretation of this predictor using matrix algebra leads to an easily interpretable form. The suggested predictor is shown to be a weighted average between the observed instantaneous growth rate and

the predicted instantaneous growth rate. Interpretation of θ is discussed following the mathematical derivation of β and \mathbf{e} .

Harville (1976) gave an extension of the Gauss-Markov theorem for prediction of random terms. Using theorem 1 from Harville (1976), the best linear unbiased predictor of \mathbf{e} can be found. This leads to θ , the individual growth rate. The random parameter θ is the sum of a fixed parameter, $\mathbf{X}\beta$ and a random term, \mathbf{e} (see (3.29)). The BLUP of θ will be the sum of the BLUE of $\mathbf{X}\beta$ and the BLUP of \mathbf{e} . From the full model (3.32), the BLUE of β is the GLS estimate

$$\hat{\beta} = [\mathbf{X}'\mathbf{F}'\mathbf{\Psi}^{-1}\mathbf{F}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{F}'\mathbf{\Psi}^{-1}\mathbf{Y} \quad (3.34)$$

which leads to $\mathbf{X}\hat{\beta}$, the BLUE of $\mathbf{X}\beta$.

Following Harville (1976), the BLUP of \mathbf{e} is

$$\tilde{\mathbf{e}} = \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Psi}^{-1}(\mathbf{Y} - \mathbf{F}\mathbf{X}\hat{\beta}) \quad (3.35)$$

which can be developed further using the definition of $\mathbf{\Psi}$ and matrix inversion rules

$$\begin{aligned} \tilde{\mathbf{e}} &= \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Psi}^{-1}\mathbf{Y} - \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Psi}^{-1}\mathbf{F}\mathbf{X}\hat{\beta} \\ &= \phi^2\mathbf{V}\mathbf{F}'\{\mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\mathbf{F}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}) + \phi^{-2}\mathbf{V}^{-1}]^{-1}\mathbf{F}'\mathbf{\Omega}^{-1}\}\mathbf{Y} \\ &\quad - \phi^2\mathbf{V}\mathbf{F}'\{\mathbf{\Omega}^{-1} - \mathbf{\Omega}^{-1}\mathbf{F}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}) + \phi^{-2}\mathbf{V}^{-1}]^{-1}\mathbf{F}'\mathbf{\Omega}^{-1}\}\mathbf{F}\mathbf{X}\hat{\beta} \\ &= \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{Y} - \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}) + \phi^{-2}\mathbf{V}^{-1}]^{-1}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{Y} \\ &\quad - \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}\mathbf{X}\hat{\beta} + \phi^2\mathbf{V}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}) + \phi^{-2}\mathbf{V}^{-1}]^{-1}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}\mathbf{X}\hat{\beta} \\ &= \phi^2\mathbf{V}\{\mathbf{I} - (\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}) + \phi^{-2}\mathbf{V}^{-1}]^{-1}\}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{Y} \\ &\quad - \phi^2\mathbf{V}\{\mathbf{I} - (\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}) + \phi^{-2}\mathbf{V}^{-1}]^{-1}\}\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}\mathbf{X}\hat{\beta} \\ &= \phi^2\mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2\mathbf{V}]^{-1}(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1}\mathbf{F}\mathbf{\Omega}^{-1}\mathbf{Y} \end{aligned}$$

$$-\phi^2 \mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1} \mathbf{X}\hat{\beta} \quad (3.36)$$

This predictor is unbiased in the sense that $E[\tilde{\mathbf{e}}] = E[\mathbf{e}] = \mathbf{0}$ where $\mathbf{0}$ is a column-vector of zeros. It is best in the sense that

$$E[(\tilde{\mathbf{e}} - \mathbf{e})'(\tilde{\mathbf{e}} - \mathbf{e})] < E[(\hat{\mathbf{e}} - \mathbf{e})'(\hat{\mathbf{e}} - \mathbf{e})] \quad (3.37)$$

where $\hat{\mathbf{e}}$ is any other predictor of \mathbf{e} .

The BLUP of θ can now be defined as

$$\tilde{\theta} = \mathbf{X}\hat{\beta} + \tilde{\mathbf{e}} \quad (3.38)$$

$$\begin{aligned} &= \mathbf{X}\hat{\beta} + \phi^2 \mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1} (\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} \mathbf{F}'\mathbf{\Omega}^{-1} \mathbf{Y} \\ &\quad - \phi^2 \mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1} \mathbf{X}\hat{\beta} \\ &= \phi^2 \mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1} (\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} \mathbf{F}'\mathbf{\Omega}^{-1} \mathbf{Y} \\ &\quad + \{\mathbf{I} - \phi^2 \mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1}\} \mathbf{X}\hat{\beta} \\ &= \phi^2 \mathbf{V}[(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1} (\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} \mathbf{F}'\mathbf{\Omega}^{-1} \mathbf{Y} \\ &\quad + (\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} [(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} + \phi^2 \mathbf{V}]^{-1} \mathbf{X}\hat{\beta} \end{aligned} \quad (3.39)$$

This predictor of θ can further be simplified; however, the terms it includes first must be recognized. The term $(\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F})^{-1} \mathbf{F}'\mathbf{\Omega}^{-1} \mathbf{Y}$ is the GLS estimate of θ , $\hat{\theta}$, from the first-stage model, model (3.18). Its distribution is conditional on θ since θ , a random variable, was assumed fixed in the first stage. This distribution is

$$\hat{\theta} | \theta \sim N(\theta, [\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}]^{-1}). \quad (3.40)$$

Therefore, the term $[\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}]^{-1}$ is the variance of the variable $\hat{\theta} | \theta$. From the distribution of θ , shown in (3.30), the variance of θ is $\phi^2 \mathbf{V}$, and its expected value is $\mathbf{X}\beta$. Combining

this information with the distribution shown in (3.40), the unconditional distribution of $\hat{\theta}$ is:

$$\hat{\theta} \sim N(\mathbf{X}\beta, \Phi) \quad (3.41)$$

where $\Phi = [\mathbf{F}'\Omega^{-1}\mathbf{F}]^{-1} + \phi^2\mathbf{V}$.

The GLS estimate of β from the full model, defined in (3.34), is equivalent to the GLS estimate of β from the unconditional distribution of $\hat{\theta}$, shown in (3.41) (Amemiya 1971). This can be shown using the matrix inversion rules.

$$\begin{aligned} \hat{\beta} &= [\mathbf{X}'\mathbf{F}'\Psi^{-1}\mathbf{F}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{F}'\Psi^{-1}\mathbf{Y} \quad (3.42) \\ &= [\mathbf{X}'\mathbf{F}'\Omega^{-1}\mathbf{F}\mathbf{X} - \mathbf{X}'\mathbf{F}'\Omega^{-1}\mathbf{F}(\mathbf{F}'\Omega^{-1}\mathbf{F} + \phi^{-2}\mathbf{V}^{-1})^{-1}\mathbf{F}'\Omega^{-1}\mathbf{F}\mathbf{X}]^{-1} \\ &\quad \cdot [\mathbf{X}'\mathbf{F}'\Omega^{-1}\mathbf{Y} - \mathbf{X}'\mathbf{F}'\Omega^{-1}\mathbf{F}(\mathbf{F}'\Omega^{-1}\mathbf{F} + \phi^{-2}\mathbf{V}^{-1})^{-1}\mathbf{F}'\Omega^{-1}\mathbf{Y}] \\ &= [\mathbf{X}'\{\mathbf{I} - (\mathbf{F}'\Omega^{-1}\mathbf{F})(\mathbf{F}'\Omega^{-1}\mathbf{F} + \phi^{-2}\mathbf{V}^{-1})^{-1}\}(\mathbf{F}'\Omega^{-1}\mathbf{F})\mathbf{X}]^{-1} \\ &\quad \cdot [\mathbf{X}'\{\mathbf{I} - (\mathbf{F}'\Omega^{-1}\mathbf{F})(\mathbf{F}'\Omega^{-1}\mathbf{F} + \phi^{-2}\mathbf{V}^{-1})^{-1}\}\mathbf{F}'\Omega^{-1}\mathbf{Y}] \\ &= [\mathbf{X}'[(\mathbf{F}'\Omega^{-1}\mathbf{F})^{-1} + \phi^2\mathbf{V}]^{-1}\mathbf{X}]^{-1} \\ &\quad \cdot \mathbf{X}'[(\mathbf{F}'\Omega^{-1}\mathbf{F})^{-1} + \phi^2\mathbf{V}]^{-1}(\mathbf{F}'\Omega^{-1}\mathbf{F})^{-1}\mathbf{F}'\Omega^{-1}\mathbf{Y} \\ &= [\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\hat{\theta} \quad (3.43) \end{aligned}$$

This demonstrates that $\mathbf{X}\hat{\beta}$ in $\tilde{\theta}$ (see model (3.39)) is equivalent to the GLS estimate of the expected value of the unconditional distribution of $\hat{\theta}$.

The BLUP of θ , equation (3.39), can therefore be rewritten:

$$\tilde{\theta} = \text{var}(\theta)[\text{var}(\hat{\theta}|\theta) + \text{var}(\theta)]^{-1}\hat{\theta} + \text{var}(\hat{\theta}|\theta)[\text{var}(\hat{\theta}|\theta) + \text{var}(\theta)]^{-1}\hat{E}[\hat{\theta}] \quad (3.44)$$

It is almost identical to the expected value of the conditional variable θ given $\hat{\theta}$. Using Bayes' rule, the expected value of $\theta|\hat{\theta}$ is shown to be:

$$E[\theta|\hat{\theta}] = \text{var}(\theta)[\text{var}(\hat{\theta}|\theta) + \text{var}(\theta)]^{-1}\hat{\theta} + \text{var}(\hat{\theta}|\theta)[\text{var}(\hat{\theta}|\theta) + \text{var}(\theta)]^{-1}E[\theta] \quad (3.45)$$

The only difference is that the unknown expected value of the unconditional variable θ is replaced by an estimate in the BLUP. The BLUP can be interpreted as a weighted average between the growth rate observed on a plot and the growth rate predicted from some plot attributes. This can easily be understood. The individual growth rates can be predicted from some independent variables with some error, and it can also be observed with some error. The best predictor will be an average between the prediction and the observation with the weights proportional to the confidence in each estimate. This confidence is inversely proportional to the respective variance of each estimate. The BLUP is this weighted average.

As mentioned before, the BLUP is a random variable. The precision of the BLUP can be obtained by estimating the variance of the predictor. This variance estimator follows.

3.3.1.2 Estimation of the Variance of the Predictor

The BLUP is a random variable, and estimation of its variance will be needed to build a confidence interval around this predictor. The individual BLUP of θ_i can be defined by:

$$\tilde{\theta}_i = \mathbf{x}_i\hat{\beta} + \tilde{e}_i \quad (3.46)$$

where $\mathbf{x}_i = [SI_i/(V_i^0)^{1/3} - SI_i]$. Because $\tilde{\theta}_i$ is a predictor of θ_i , the variance of $\tilde{\theta}_i$ can be expressed as:

$$\begin{aligned} E[(\tilde{\theta}_i - \theta_i)^2] &= E[(\mathbf{x}_i\hat{\beta} + \tilde{e}_i - \mathbf{x}_i\beta - e_i)^2] \\ &= E[((\mathbf{x}_i\hat{\beta} - \mathbf{x}_i\beta) + \tilde{e}_i - e_i)^2] \end{aligned}$$

$$\begin{aligned}
&= E[(\mathbf{x}_i\hat{\beta} - \mathbf{x}_i\beta)^2] + E[\tilde{e}_i^2] + E[e_i^2] + 2E[(\mathbf{x}_i\hat{\beta} - \mathbf{x}_i\beta)\tilde{e}_i] \\
&\quad - 2E[(\mathbf{x}_i\hat{\beta} - \mathbf{x}_i\beta)e_i] - 2E[\tilde{e}_ie_i] \\
&= \text{var}(\mathbf{x}_i\hat{\beta}) + \text{var}(\tilde{e}_i) + \text{var}(e_i) + 2\text{cov}(\mathbf{x}_i\hat{\beta}, \tilde{e}_i) - 2\text{cov}(\tilde{e}_i, e_i) \quad (3.47)
\end{aligned}$$

The various components can be estimated separately, using the mathematics of linear combinations. First, it can be shown that the covariance between $\hat{\beta}$ and \tilde{e} is zero which means that $\text{cov}(\mathbf{x}_i\hat{\beta}, \tilde{e}_i)$ is zero.

$$\begin{aligned}
\text{cov}(\hat{\beta}, \tilde{e}) &= \text{cov}([\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\hat{\theta}, \phi^2\mathbf{V}\Phi^{-1}[\hat{\theta} - \mathbf{X}\hat{\beta}]) \\
&= \text{cov}([\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\hat{\theta}, \{\phi^2\mathbf{V}\Phi^{-1}[\mathbf{I} - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}]\}\hat{\theta}) \\
&= [\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\text{var}(\hat{\theta})\{\phi^2\mathbf{V}\Phi^{-1}[\mathbf{I} - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}]\}' \\
&= [\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\Phi\{[\mathbf{I} - \Phi^{-1}\mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}']\Phi^{-1}\phi^2\mathbf{V}\} \\
&= [\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'[\Phi^{-1}\phi^2\mathbf{V} - \Phi^{-1}\mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\phi^2\mathbf{V}] \\
&= [\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\phi^2\mathbf{V} - [\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\phi^2\mathbf{V} \\
&= \mathbf{0} \quad (3.48)
\end{aligned}$$

The variance of the BLUP of \tilde{e}_i can be derived from the definition of \tilde{e} in (3.36). Let $\{\phi^2\mathbf{V}\Phi^{-1}\}_i$ be the i^{th} row of the matrix defined by $\phi^2\mathbf{V}\Phi^{-1}$. Then the variance of \tilde{e}_i is:

$$\begin{aligned}
\text{var}(\tilde{e}_i) &= \text{var}(\{\phi^2\mathbf{V}\Phi^{-1}\}_i[\hat{\theta} - \mathbf{X}\hat{\beta}]) \\
&= \{\phi^2\mathbf{V}\Phi^{-1}\}_i\text{var}(\hat{\theta} - \mathbf{X}\hat{\beta})\{\phi^2\mathbf{V}\Phi^{-1}\}_i' \\
&= \{\phi^2\mathbf{V}\Phi^{-1}\}_i\text{var}([\mathbf{I} - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}]\hat{\theta})\{\phi^2\mathbf{V}\Phi^{-1}\}_i' \\
&= \{\phi^2\mathbf{V}\Phi^{-1}\}_i[\mathbf{I} - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}]\text{var}(\hat{\theta}) \\
&\quad \cdot [\mathbf{I} - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}]'\{\phi^2\mathbf{V}\Phi^{-1}\}_i' \\
&= \{\phi^2\mathbf{V}\Phi^{-1}\}_i[\mathbf{I} - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}]\Phi
\end{aligned}$$

$$\begin{aligned}
& [\mathbf{I} - \Phi^{-1}\mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}']\{\phi^2\mathbf{V}\Phi^{-1}\}_i' \\
&= \{\phi^2\mathbf{V}\Phi^{-1}\}_i[\Phi - \mathbf{X}[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}']\{\phi^2\mathbf{V}\Phi^{-1}\}_i'
\end{aligned} \tag{3.49}$$

To define the covariance between $\tilde{\theta}_i$ and e_i , it will be useful to express $\tilde{\theta}_i$ in terms of \mathbf{Y} . From equation (3.38), $\tilde{\theta}_i$ can be written as:

$$\begin{aligned}
\tilde{\theta}_i &= \mathbf{x}_i\hat{\beta} + \{\phi^2\mathbf{V}\Phi^{-1}\}_i[\hat{\theta} - \mathbf{X}\hat{\beta}] \\
&= [\mathbf{x}_i - \{\phi^2\mathbf{V}\Phi^{-1}\}_i\mathbf{X}]\hat{\beta} + \{\phi^2\mathbf{V}\Phi^{-1}\}_i\hat{\theta} \\
&= [\mathbf{x}_i - \{\phi^2\mathbf{V}\Phi^{-1}\}_i\mathbf{X}][\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1}\hat{\theta} + \{\phi^2\mathbf{V}\Phi^{-1}\}_i\hat{\theta} \\
&= ([\mathbf{x}_i - \{\phi^2\mathbf{V}\Phi^{-1}\}_i\mathbf{X}][\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1} + \{\phi^2\mathbf{V}\Phi^{-1}\}_i)(\mathbf{F}'\Omega^{-1}\mathbf{F})^{-1}\mathbf{F}'\Omega^{-1}\mathbf{Y} \\
&= \Gamma_i\mathbf{Y}
\end{aligned} \tag{3.50}$$

where $\Gamma_i = ([\mathbf{x}_i - \{\phi^2\mathbf{V}\Phi^{-1}\}_i\mathbf{X}][\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Phi^{-1} + \{\phi^2\mathbf{V}\Phi^{-1}\}_i)(\mathbf{F}'\Omega^{-1}\mathbf{F})^{-1}\mathbf{F}'\Omega^{-1}$.

Therefore, the covariance of $\tilde{\theta}_i$ and e_i is

$$\begin{aligned}
\text{cov}(\tilde{\theta}_i, e_i) &= \text{cov}(\Gamma_i\mathbf{Y}, e_i) \\
&= \text{cov}(\Gamma_i(\mathbf{F}\mathbf{X}\beta + \mathbf{F}\mathbf{e} + \mathbf{G}\mathbf{r} + \mathbf{v}), e_i) \\
&= \text{cov}(\Gamma_i\mathbf{F}\mathbf{e}, e_i) \\
&= \text{cov}(\Gamma_i\mathbf{F}_i e_i, e_i) \\
&= \Gamma_i\mathbf{F}_i \text{var}(e_i)
\end{aligned} \tag{3.51}$$

The variance of e_i is given by

$$\text{var}(e_i) = \phi^2/(V_i^0)^{1/3} \tag{3.52}$$

The term $\text{var}(\mathbf{x}_i\hat{\beta})$ is easily obtained from the variance of $\hat{\beta}$:

$$\begin{aligned}
\text{var}(\mathbf{x}_i\hat{\beta}) &= \mathbf{x}_i \text{var}(\hat{\beta}) \mathbf{x}_i' \\
&= \mathbf{x}_i[\mathbf{X}'\Phi^{-1}\mathbf{X}]^{-1}\mathbf{x}_i'
\end{aligned} \tag{3.53}$$

Using these different estimators, the variance of $\tilde{\theta}_i$ can then be defined as

$$\begin{aligned} E[(\tilde{\theta}_i - \theta_i)^2] &= \mathbf{x}_i[\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}]^{-1}\mathbf{x}_i' + \{\phi^2\mathbf{V}\mathbf{\Phi}^{-1}\}_i[\mathbf{\Phi} - \mathbf{X}[\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}]^{-1}\mathbf{X}']\{\phi^2\mathbf{V}\mathbf{\Phi}^{-1}\}_i' \\ &\quad + (1 - 2\mathbf{\Gamma}_i\mathbf{F}_i)\phi^2/(V_i^0)^{1/3} \end{aligned} \quad (3.54)$$

This estimator yields an estimate of the variance of the predictor of θ_i . Both, the predictor of θ_i and the estimator of the predictor's variance are based on the assumption that the covariance matrix of the error terms \mathbf{e} , \mathbf{r} , and \mathbf{v} are known. These unknown matrices must now be replaced by their estimates.

3.3.2 Covariance Matrix Estimators

Knowing the predictor when the covariance matrices are known “may provide some insights into what to do when [they are] not” (Harville 1976). A common approach is to replace the unknown covariance matrices by their estimates and to assume that they are the true values. The various assumptions about the error terms have drastically reduced the number of unknown parameters. Estimates of δ^2 , $\sigma_1^2, \dots, \sigma_M^2$, and ϕ^2 are sufficient to obtain estimates of the matrices $\mathbf{\Omega}$ and $\phi^2\mathbf{V}$.

The best strategy for estimating the unknown variances is to use the two-stage structure of the model; each stage can be dealt with independently. In the first stage, estimates of θ , δ^2 and $\sigma_1^2, \dots, \sigma_M^2$ based on the observed growth rates are obtained. In the second stage, the variable θ is unobservable, but the unconditional distribution of $\hat{\theta}$, an estimate of θ from the first stage, can be derived. From this distribution, estimates of β and ϕ^2 are computed.

3.3.2.1 First-Stage Estimators

Different estimators for error component models have been suggested in the literature. However, none of these estimators were found adequate for the assumptions made in the

first stage. Instead, theoretical estimators had to be derived based on moment estimators of the unobserved residuals. It is a common practice to replace the unobserved residuals by their OLS estimates in theoretical estimators (Wallace and Hussain 1969, Zellner 1962, Swamy 1971, among others). For small sample sizes, when replacing the unobserved residuals by their OLS estimates in an estimator, it is probably worthwhile to adjust the degrees of freedom.

A) Estimator of σ_p^2

Following Wallace and Hussain (1969), the estimator for σ_p^2 can be derived using the squared residuals from period p with the following reasoning:

$$\begin{aligned}
 E \left[\sum_i \left[u_{ip} - \frac{\sum_i u_{ip}}{S_p} \right]^2 \right] &= E \left[\sum_i \left[r_p + v_{ip} - \frac{(S_p r_p + \sum_i v_{ip})}{S_p} \right]^2 \right] \\
 &= E \left[\sum_i \left[v_{ip} - \frac{\sum_i v_{ip}}{S_p} \right]^2 \right] \\
 &= E \left[\sum_i [v_{ip} - \bar{v}_p]^2 \right] \\
 &= \sum_i E [v_{ip} - \bar{v}_p]^2 \\
 &= S_p \sigma_p^2
 \end{aligned} \tag{3.55}$$

Therefore, the theoretical estimator for σ_p^2 can be

$$\hat{\sigma}_p^2 = \frac{1}{S_p} \sum_i \left[u_{ip} - \frac{\sum_i u_{ip}}{S_p} \right]^2 \tag{3.56}$$

which, after correcting for the degrees of freedom and replacing the unknown residuals by their OLS estimates \hat{u}_{ip} , becomes

$$\hat{\sigma}_p^2 = \frac{1}{(S_p - 1)} \sum_i \left[\hat{u}_{ip} - \frac{\sum_i \hat{u}_{ip}}{S_p} \right]^2 \tag{3.57}$$

B) Estimator of δ^2

Using a similar strategy as for $\hat{\sigma}_p^2$, an estimator for δ^2 can be derived using the squared residuals, the estimates of σ_p^2 ($p = 1, \dots, M$), the common variance of time-dependent error terms coming from the same time period and the covariance between error terms from overlapping time periods.

$$E \left[\sum_p \left(\sum_i u_{ip}^2 - S_p \sigma_p^2 \right) \right] = W \delta^2 \quad (3.58)$$

$$\sum_p E \left[\sum_i \sum_{j>i} u_{ip} u_{jp} \right] = \delta^2 \sum_p \frac{(S_p^2 - S_p)}{2} \quad (3.59)$$

$$E \left[\sum_i \sum_p \sum_j \sum_{q>p} u_{ip} u_{jq} \right] = \delta^2 \sum_p \sum_{p>q} S_p S_q \lambda_{pq} \quad (3.60)$$

From the sum of these three estimators, the estimator of δ^2 is easily obtained

$$\hat{\delta}^2 = \frac{1}{\nu} \left(\sum_i \sum_p \sum_j \sum_q u_{ip} u_{jq} - \sum_p S_p \sigma_p^2 \right) \quad (3.61)$$

where $\nu = W + \sum_p \frac{(S_p^2 - S_p)}{2} + \sum_p \sum_{q>p} S_p S_q \lambda_{pq}$.

After correcting for the degrees of freedom and replacing the unknown residuals by their OLS residuals, the estimator becomes

$$\hat{\delta}^2 = \frac{1}{(\nu - 1)} \left(\sum_i \sum_p \sum_j \sum_{\substack{q \\ (q \cap p) \neq \emptyset}} \hat{u}_{ip} \hat{u}_{jq} - \sum_p S_p \hat{\sigma}_p^2 \right) \quad (3.62)$$

C) Estimator of $\hat{\theta}$

The estimates of σ_p^2 and δ^2 are replaced in Ω to get the estimate $\hat{\Omega}$. The FGLS estimate of θ becomes:

$$\hat{\hat{\theta}} = [\mathbf{F}' \hat{\Omega}^{-1} \mathbf{F}]^{-1} \mathbf{F}' \hat{\Omega}^{-1} \mathbf{Y} \quad (3.63)$$

Again, the double hat above θ implies that it was obtained after first estimating Ω . The FGLS estimate is asymptotically normally distributed. When W is large, the normal

distribution should be a relatively good approximation. The approximate distribution of $\hat{\theta}|\theta$, is:

$$\hat{\theta}|\theta \sim N(\theta, [\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F}]^{-1}) \quad (3.64)$$

The inversion of matrix $\hat{\Omega}$, a $W \times W$ matrix, can be simplified using matrix inversion rules. The inversion of $\hat{\Omega}$ is equivalent to:

$$\begin{aligned} \hat{\Omega}^{-1} &= [\delta^2 \mathbf{G} \mathbf{\Lambda} \mathbf{G}' + \hat{\Sigma}]^{-1} \\ &= \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} \mathbf{G} \left[\mathbf{G}' \hat{\Sigma}^{-1} \mathbf{G} + 1/\delta^2 \mathbf{\Lambda}^{-1} \right]^{-1} \mathbf{G}' \hat{\Sigma}^{-1} \end{aligned} \quad (3.65)$$

which requires the inversion of $\hat{\Sigma}$, a diagonal matrix, and $\mathbf{\Lambda}$ and $\left[\mathbf{G}' \hat{\Sigma}^{-1} \mathbf{G} + 1/\delta^2 \mathbf{\Lambda}^{-1} \right]$, two $M \times M$ matrices.

3.3.2.2 Second-Stage Estimators

In the second stage, an estimate of the unconditional distribution of $\hat{\theta}$, the GLS estimate of θ is needed. Unfortunately, the GLS estimate of θ cannot be obtained because the covariance matrix of the error terms is unknown. This difficulty can be overcome using the unconditional distribution of $\hat{\theta}$, a FGLS estimate of θ . From the conditional distribution of $\hat{\theta}$ given θ , shown in (3.64), and the distribution of θ , shown in (3.30), the unconditional distribution of $\hat{\theta}$ is easily found:

$$\hat{\theta} \sim N(\mathbf{X}\beta, [\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F}]^{-1} + \phi^2 \mathbf{V}) \quad (3.66)$$

In the second-stage, estimates of β and ϕ^2 are needed. GLS estimates are not possible because they would depend on the unknown value of each other. An iterative FGLS estimate is then required.

Theoretically, the GLS estimate of β would be

$$\hat{\beta} = [\mathbf{X}'[\phi^2 \mathbf{V} + (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}]\mathbf{X}]^{-1} \mathbf{X}'[\phi^2 \mathbf{V} + (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}]^{-1} \hat{\theta} \quad (3.67)$$

which is also the maximum likelihood estimate since the error terms are assumed to be normally distributed.

Similarly, a possible estimate of ϕ^2 is its maximum likelihood estimate. However, equating the differential of the pdf of $\hat{\theta}$ with respect to ϕ^2 with zero does not yield a simple solution. No satisfactory estimate was found in the literature. An alternative estimate had to be derived.

Based on the following reasoning:

$$\begin{aligned}
 E[(\hat{\theta} - \mathbf{X}\beta)'(\hat{\theta} - \mathbf{X}\beta)] &= \text{tr}[\phi^2 \mathbf{V} + (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}] \\
 &= \phi^2 \text{tr}[\mathbf{V}] + \text{tr}[(\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}] \\
 E \left[\frac{(\hat{\theta} - \mathbf{X}\beta)'(\hat{\theta} - \mathbf{X}\beta) - \text{tr}[(\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}]}{\text{tr}[\mathbf{V}]} \right] &= \phi^2
 \end{aligned} \tag{3.68}$$

the following estimate was found adequate

$$\hat{\phi}^2 = \frac{N}{(N-2)} \frac{(\hat{\theta} - \mathbf{X}\hat{\beta})'(\hat{\theta} - \mathbf{X}\hat{\beta}) - \text{tr}[(\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}]}{\text{tr}[\mathbf{V}]} \tag{3.69}$$

Because the unknown β is replaced by an estimate, 2 degrees of freedom are lost for the two coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$.

Since both estimates, $\hat{\phi}^2$ and $\hat{\beta}$, depend on the unknown value of the other parameter, FGLS estimates are needed. To find these FGLS estimates, an iterative solution must be used. The algorithm stops when both $\hat{\beta}$ and $\hat{\phi}^2$ converge. The iterative estimate of ϕ^2 is used to get the FGLS estimate of β :

$$\hat{\beta} = [\mathbf{X}'[\hat{\phi}^2 \mathbf{V} + (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}]^{-1} \mathbf{X}]^{-1} \mathbf{X}'[\hat{\phi}^2 \mathbf{V} + (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}]^{-1} \hat{\theta} \tag{3.70}$$

3.3.3 The Best Linear Unbiased Predictor and Its Variance

The estimates of σ_p^2 , δ^2 , ϕ^2 , and β can now be replaced in the BLUP, defined in equation (3.39), to obtain what could be called an “estimated best linear unbiased predictor”. Equation (3.39) becomes:

$$\begin{aligned}\tilde{\theta} &= \hat{\phi}^2 \mathbf{V}[(\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1} + \hat{\phi}^2 \mathbf{V}]^{-1} \hat{\theta} + \\ &\quad (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1}[(\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1} + \hat{\phi}^2 \mathbf{V}]^{-1} \mathbf{X}\hat{\beta}\end{aligned}\quad (3.71)$$

Similarly, an estimate of the variance of the BLUP of θ_i is obtained by replacing the unknown parameters in (3.54) by their estimates. This yields:

$$\begin{aligned}E[(\tilde{\theta}_i - \theta_i)^2] &= \mathbf{x}_i[\mathbf{X}'\hat{\Phi}^{-1}\mathbf{X}]^{-1}\mathbf{x}_i' + \{\hat{\phi}^2 \mathbf{V}\hat{\Phi}^{-1}\}_i[\hat{\Phi} - \mathbf{X}[\mathbf{X}'\hat{\Phi}^{-1}\mathbf{X}]^{-1}\mathbf{X}']\{\hat{\phi}^2 \mathbf{V}\hat{\Phi}^{-1}\}_i' \\ &\quad + (1 - 2\hat{\Gamma}_i\mathbf{F}_i)\hat{\phi}^2/(V_i^0)^{1/3}\end{aligned}\quad (3.72)$$

where

$$\begin{aligned}\hat{\Phi} &= (\mathbf{F}'\hat{\Omega}^{-1}\mathbf{F})^{-1} + \hat{\phi}^2 \mathbf{V} \\ \hat{\Gamma}_i &= ([\mathbf{x}_i - \{\hat{\phi}^2 \mathbf{V}\hat{\Phi}^{-1}\}_i\mathbf{X}][\mathbf{X}'\hat{\Phi}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\hat{\Phi}^{-1} + \{\hat{\phi}^2 \mathbf{V}\hat{\Phi}^{-1}\}_i)(\mathbf{F}\hat{\Omega}^{-1}\mathbf{F})^{-1}\mathbf{F}'\hat{\Omega}^{-1}\end{aligned}$$

The estimate of the BLUP of θ_i can now be used to predict future growth rates and compound the current volume of a permanent sample plot. Confidence intervals around the prediction can be built with the variance estimate.

3.4 Model Predictions

The annual growth rate on a given plot can be predicted from model (3.46)

$$\tilde{\theta}_i = \mathbf{x}_i\hat{\beta} + \tilde{e}_i \quad (3.73)$$

Once the unknown parameters β and e_i have been estimated, the instantaneous growth rate can be predicted from two plot attributes: the site index and the initial

volume. The growth rate can be assumed to be constant in a given year and to be changing from year to year. The volume can thus be updated annually. Each updated volume becomes the new initial volume, (*i.e.* the volume at the beginning of the next growing season). The growth rate that will occur in this particular year is then predicted leading to a new updated volume. This process is repeated, say, K times to get the volume K years from now. This recursive system offers more flexibility to predict future volumes because the growth rate is updated annually. Assuming a constant instantaneous growth rate over a period of one year is close to reality.

The annual volume update involves computing the expected value of the antilog of the predictor $\tilde{\theta}_i$. The technique for estimating this expected value is discussed in section 3.4.1. If a confidence interval is desired around the predicted volume, the confidence interval around each intermediate volume will be needed too since the volume at the end of the growing season is a function of the predicted volume at the beginning of the season. In section 3.4.2, the problem of variance estimation is analyzed, and a solution is suggested.

3.4.1 Volume Predictor

Various techniques exist to estimate the expected value of the antilog of an estimate (Ung and Végiaard 1988). However, no technique was found to estimate the expected value of the antilog of a predictor. The subtle difference can be explained by recalling that $\tilde{\theta}_i$ is a predictor of θ_i , a random variable, while the expected value of an estimate is a constant. The technique used by Baskerville (1972) was simple enough to be adapted for estimating the expected value of the antilog of a predictor and was, consequently, the chosen technique.

Future volume prediction is driven by the initial volume in the first iteration. The last volume measurement observed on the sample plot can serve as the initial volume since

this volume represents the best knowledge of initial volume to predict future growth rates.

A possible estimator of the updated volume could be the volume at the beginning of the growing season multiplied by the antilog of the predicted growth rate:

$$\hat{V}_i^{j+1} = \hat{V}_i^j \times \exp[\tilde{\theta}_i^j] \quad (3.74)$$

where \hat{V}_i^j is the predicted volume on plot i at iteration j and $\tilde{\theta}_i^j$ is the predicted growth rate. However, this estimator is biased. Based on the moment generating function, it can be shown that

$$E[X] = \exp \left[E[\ln X] + \frac{\text{var}(\ln X)}{2} \right] \quad (3.75)$$

which can be approximated by

$$\hat{E}[X] = \exp \left[\hat{E}[\ln X] + \frac{\widehat{\text{var}}(\ln X)}{2} \right] \quad (3.76)$$

Goldberger (1968) pointed out that this result was asymptotically unbiased.²

The true model expressed in a logarithmic scale is

$$\ln V_i^{j+1} = \ln V_i^j + \theta_i^j + u_i^j \quad (3.77)$$

where u_i^j is the error term on plot i at iteration j . The first term on the right-hand side of the equation can also be rewritten in a similar form. Hence, the true model really is

$$\ln V_i^{j+1} = \ln V_i^0 + \sum_{k=0}^j \theta_i^k + \sum_{k=0}^j u_i^k \quad (3.78)$$

Given that the volume is estimated on plot i and that annual observations are used, if the first volume is assumed to be measured without error, then

$$E[\ln V_i^{j+1}] = \ln V_i^0 + \sum_{k=0}^j \theta_i^k. \quad (3.79)$$

²Goldberger (1968) also mentioned that, instead of using the corrected expected value of $\ln X$, one can use the uncorrected median. However, the median might be more difficult to obtain than the corrected mean.

The variance of $\ln V_i^{j+1}$ is the variance of the sum of the error terms

$$\begin{aligned}\text{var}(\ln V_i^{j+1}) &= \text{var}\left(\sum_{k=0}^j u_i^k\right) \\ &= \sum_{k=0}^j \text{var}(u_i^k)\end{aligned}\quad (3.80)$$

The moment generating function can be used to rewrite the expected value of V_i^{j+1}

$$\begin{aligned}E[V_i^{j+1}] &= \exp\left[E[\ln V_i^{j+1}] + \frac{\text{var}(\ln V_i^{j+1})}{2}\right] \\ &= \exp\left[\ln V_i^0 + \sum_{k=0}^j \theta_i^k + \frac{\sum_{k=0}^j \text{var}(u_i^k)}{2}\right]\end{aligned}\quad (3.81)$$

Baskerville (1972) replaced the theoretical moment generating function by an estimate. The same strategy can be followed. At this point, no estimate of $\text{var}(u_i^k)$ exists because the time period which u_i^k represents is located in the future. A possible estimate is to take the average variance estimate of the past time periods. This estimate can be found using

$$s^2 = \frac{(\mathbf{Y} - \mathbf{F}\tilde{\theta})'(\mathbf{Y} - \mathbf{F}\tilde{\theta})}{W - 3}\quad (3.82)$$

where 2 degrees of freedom are lost for the estimates of β_1 and β_2 and one is lost for $\sum \hat{e}_i = 0$. Any error terms located in the future will be assumed to have this estimated variance. The estimate of $E[\hat{V}_i^{j+1}]$ will then be

$$\hat{E}[\hat{V}_i^{j+1}] = \exp\left[\ln V_i^0 + \sum_{k=0}^j \tilde{\theta}_i^k + \frac{(j+1)s^2}{2}\right]\quad (3.83)$$

This estimator is the closest form to the estimator used by Baskerville (1972). Its main advantages are that it is simple to compute and asymptotically unbiased. Estimation of a confidence interval of this predicted volume is discussed in the next section.

3.4.2 Confidence Intervals Estimator

Confidence interval estimation is also a problem when a scale transformation is needed. Ung and Végiard (1988) suggested the following confidence interval for $\hat{E}[\hat{V}_i^{j+1}]$ when the

estimated moment generating function is used

$$\left[\exp \left[a + \frac{\widehat{\text{var}}(\ln X)}{2} \right], \exp \left[b + \frac{\widehat{\text{var}}(\ln X)}{2} \right] \right] \quad (3.84)$$

where

$$a = \hat{E}[\ln X] - t(1 - \frac{\alpha}{2}; \nu) \widehat{\text{var}}(\hat{E}[\ln X])$$

$$b = \hat{E}[\ln X] + t(1 - \frac{\alpha}{2}; \nu) \widehat{\text{var}}(\hat{E}[\ln X])$$

$$\alpha = \text{type I error allowed}$$

$$\nu = \text{number of degrees of freedom}$$

A possible estimator for the variance of the estimated expected value of $\ln V_i^{j+1}$ is

$$\begin{aligned} \widehat{\text{var}}(\hat{E}[\ln V_i^{j+1}]) &= \widehat{\text{var}}(\ln V_i^0 + \sum_{k=0}^j \tilde{\theta}_i^k) \\ &= \sum_{k=0}^j \sum_{l=0}^j \widehat{\text{cov}}(\tilde{\theta}_i^k, \tilde{\theta}_i^l) \end{aligned} \quad (3.85)$$

where the estimated covariance between $\tilde{\theta}_i^k$ and $\tilde{\theta}_i^l$ is approximated by the following

$$\begin{aligned} \hat{E}[(\tilde{\theta}_i^k - \theta_i^k)(\tilde{\theta}_i^l - \theta_i^l)] &\cong \hat{\mathbf{x}}_i^k [\mathbf{X}' \hat{\mathbf{\Phi}}^{-1} \mathbf{X}]^{-1} \hat{\mathbf{x}}_i^{l'} + \\ &\quad \{ \hat{\phi}^2 \mathbf{V} \hat{\mathbf{\Phi}}^{-1} \}_i [\hat{\mathbf{\Phi}} - \mathbf{X} [\mathbf{X}' \hat{\mathbf{\Phi}}^{-1} \mathbf{X}]^{-1} \mathbf{X}'] \{ \hat{\phi}^2 \mathbf{V} \hat{\mathbf{\Phi}}^{-1} \}'_i \\ &\quad + 1 - \hat{\Gamma}_i^k \mathbf{F}_i \hat{\phi}^2 / (\hat{V}_i^k)^{1/3} \hat{\Gamma}_i^l \mathbf{F}_i \hat{\phi}^2 / (\hat{V}_i^l)^{1/3} \end{aligned} \quad (3.86)$$

where $\hat{\mathbf{x}}_i^k$ and $\hat{\Gamma}_i^k$ are the respective estimates of \mathbf{x}_i and Γ_i after the k^{th} iteration.

Following Ung and Végiard (1988) and assuming a large number of plots, the confidence interval can be estimated approximately with the following bounds

$$\begin{aligned} \text{lower bound} &= \exp \left[\ln V_i^0 + \sum_{k=0}^j \tilde{\theta}_i^k - z_{1-\frac{\alpha}{2}} \widehat{\text{cov}}(\tilde{\theta}_i^k, \tilde{\theta}_i^l) + \frac{(j+1)s^2}{2} \right], \\ \text{upper bound} &= \exp \left[\ln V_i^0 + \sum_{k=0}^j \tilde{\theta}_i^k + z_{1-\frac{\alpha}{2}} \widehat{\text{cov}}(\tilde{\theta}_i^k, \tilde{\theta}_i^l) + \frac{(j+1)s^2}{2} \right], \end{aligned} \quad (3.87)$$

This estimator is an approximation of the true variance of the predicted volume due to the extra variation created by replacing the fixed \mathbf{x}_i^k , \mathbf{x}_i^l , Γ_i^k , and Γ_i^l with estimates. Since the true variances of the various random terms in the model were also replaced by estimates, the true variance of the predicted volume will likely be underestimated.

Chapter 4

Numerical Example

4.1 Description of the Data Base

The data base was kindly provided by MacMillan Bloedel Ltd. It consisted of 68 permanent sample plots of coastal Douglas-fir. The sample plots were fixed plots located in central Vancouver Island. These were pure, even-aged, and undisturbed plots. A sample plot was considered pure Douglas-fir when Douglas-fir represented more than 80% of the basal area. Undisturbed meant that there had been no human intervention to modify the growth rate, nor catastrophic losses due to biotic or climatic agents. All sample plots were measured after the growing season. They contained between 3 and 8 measurements covering a wide range of site index (Figure 4.2). Site index was estimated with King's (1966) site index equation. To protect the confidentiality of the data, the actual volumes were multiplied by a constant factor. The modified volumes were assumed to be the measured volumes, and no further reference in the text will be made to the unmodified volumes. At first measurement, volume varied between 10.4 and 939.5 m³/ha, and breast height age varied between 8 and 90 years. At last measurement, volume varied between 95.8 and 1594.4 m³/ha, and breast height age varied between 24 and 115 years (Figure 4.3).

The initial data base was split into two distinct groups: an experimental set and a prediction set. The model was calibrated with the experimental set in order to predict the volumes in the prediction set. All measurements prior to or taken in 1972 were



Figure 4.2: Distribution of Site Index

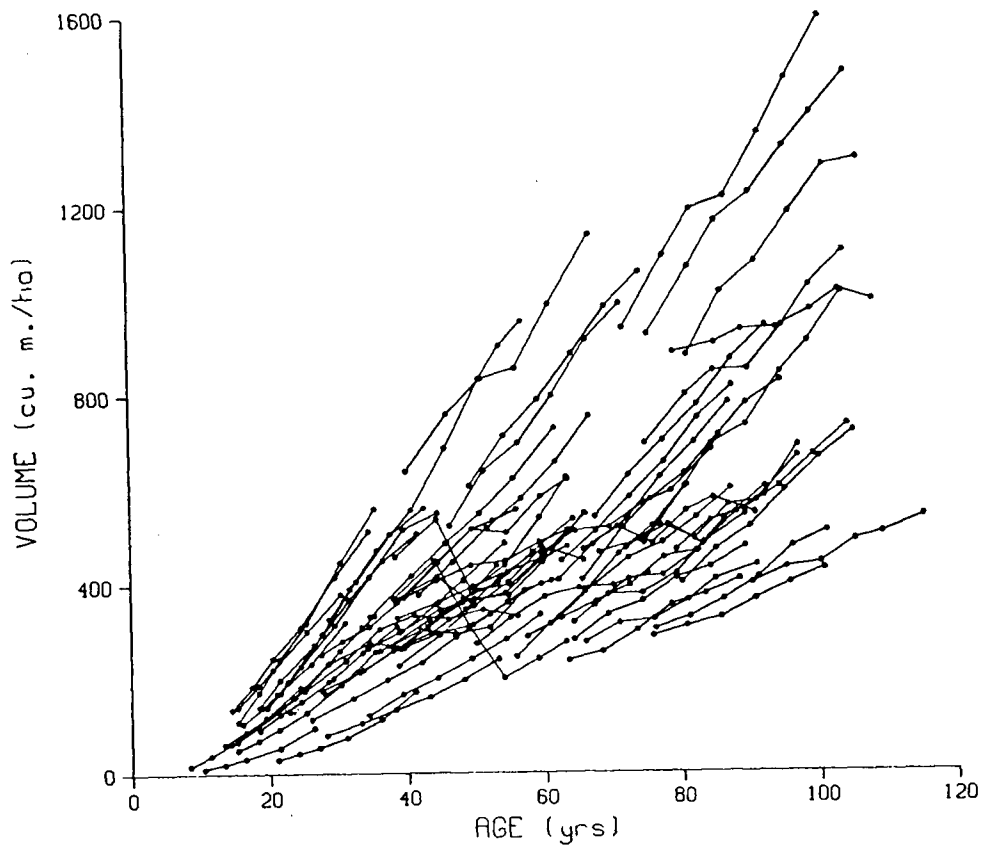


Figure 4.3: Individual Plot Volume versus Age

included in the experimental set, and the measurements taken after 1972 were put into the prediction set.

In the experimental data set, 42 plots were measured three times, and 26 plots had four measurements. In this data set, the maximum age was 100 years, and the maximum volume was 1224.6 m³/ha. The mean volume was 393.7 m³/ha and the standard deviation 244.3 m³/ha. The number of growth observations on a given plot was equal to the number of volume measurements on this particular plot minus one. For example, three measurements on a plot will yield two growth observations. Consequently, 42 plots had 2 growth observations ($T_i = 2$) and 26 plots had 3 ($T_i = 3$) for a total of 162 growth observations ($W = 162$). The growth observations in the experimental set covered 15 time intervals ($M = 15$) of 5 years on average, a minimum of 2 years and a maximum of 6 years. The number of observations in a time period varied between 2 and 26 (Table 4.1).

Table 4.1: Time Intervals

p	Interval	S_p	p	Interval	S_p
1	1955–1961	17	9	1964–1969	14
2	1956–1962	7	10	1965–1970	17
3	1959–1962	2	11	1966–1969	9
4	1959–1964	14	12	1966–1971	26
5	1960–1965	14	13	1967–1970	2
6	1961–1966	17	14	1969–1972	9
7	1962–1965	3	15	1970–1972	2
8	1962–1966	9			

In the prediction set, there were a total of 189 volumes to be predicted. The average volume in the prediction set was 555.4 m³/ha, and the standard deviation was 289.6 m³/ha. The minimum age was 19 years and the minimum volume was 57.2 m³/ha

The time lag between a predicted volume and the last volume from the same plot used in the experimental set was called *projection time*. The model performances for volume projection were evaluated with respect to this projection time. The projection time varied between 4 and 16 years. It was divided into three classes: 5 years or less, between 6 and 10 years, and more than 10 years. The first class contained 68 volumes, the second one 62 and the last one 59 (Table 4.2). Another criterion of interest was the initial age. The initial ages were grouped by decades, except for the first category that included the first two decades. Predictive ability was estimated for each projection time-initial age class.

Table 4.2: Frequency of Volumes by Initial Age and Projection Time Class

Projection Class (yrs)	Initial Age (yrs)								Total
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	
0-5	12	11	14	4	3	12	9	3	68
6-10	12	11	8	4	3	12	9	3	62
11+	8	13	7	4	3	12	9	3	59
Total	32	35	29	12	9	36	27	9	189

It was also interesting to look at the predictive ability of the model as a function of the time lag between the predicted volume and the first observed volume on the same plot. This was called *total lag*. The total lag varied between 10 and 31 years. It was grouped in 4 categories: 10-15 years, 16-20 years, 21-25 years, and 26 years or more. The first class had 51 observations, the second one 49, the third one 66, and the last

one 23 (Table 4.3). The interaction initial age-total lag was also considered important to analyze, hence, the total lag classes were subdivided into age classes.

Table 4.3: Frequency of Volumes by Initial Age and Total Lag Class

Total Lag (yrs)	Initial Age							Total	
	0-20	21-30	31-40	41-50	51-60	61-70	71-80		81-90
10-15	20	6	1	4	3	10	4	3	51
16-20	10	7	1	4	3	12	9	3	49
21-25	2	12	21	4	3	12	9	3	66
26+	-	10	6	-	-	2	5	-	23
Total	32	35	29	12	9	36	27	9	189

4.2 Goodness of Fit Criteria

A model can be assessed with many techniques and statistics. Generally these statistics are computed for the original model, but sometimes the actual variable of interest is a transformation of the original model. In these cases, the model can be assessed for either the original or the transformed model.

Growth rate was the estimated variable in the original model. When variances are known, the suggested predictor is BLUP for the true growth rate. Best in this case meant that it minimizes the total sum of squared error of the growth rates. However, because the variable of interest was volume, it was considered more important to assess the capability of the model to predict volume rather than growth rate, even if the estimator was not minimum for the volume squared error. Therefore, it would theoretically be possible to

find another predictor of the vector of growth rates that would reduce the total sum of squared error on volume.

The model was tested for its ability to predict future volumes. It was fitted using the experimental data set. First, the observed growth rate was computed for each growth period by taking the difference in the natural logarithmic scale between the final volume and the initial volume and dividing this difference by the length of the period in years. Then, the first-stage model was fitted using PROC MATRIX in SAS (Sas Institute Inc., 1982) with the FUZZ option on the mainframe at the University of British Columbia. The FUZZ option reassigns any value less than 1×10^{-8} to zero.

Before the second-stage, the average of the initial volumes of all growth periods on each plot were computed. This defined V_i^0 . The second-stage model was fitted using the estimates of the growth rates from the first-stage, the plot site index and V_i^0 . The convergence criterion was

$$\begin{aligned} \hat{\beta}_1^{k+1} - \hat{\beta}_1^k &< 0.0000001 \\ \text{and } \hat{\beta}_2^{k+1} - \hat{\beta}_2^k &< 0.0000001 \end{aligned}$$

where $\hat{\beta}_1^k$ and $\hat{\beta}_2^k$ are the estimates of β_1 and β_2 after the k^{th} iteration.

Estimates of the unknown variances and of β and \mathbf{e} were replaced in the theoretical predictor. The last volume on a plot in the experimental data set was used as the initial volume for all the projections on this plot. The predicted volumes were then compared with the corresponding observed volumes in the prediction set.

The statistics used to assess the performance of the model were the bias, the mean absolute deviation and the root mean squared error, in both actual and relative scales. Also, these statistics were computed for both projection lag-initial age classes and total lag-initial age classes. Mathematically, these statistics were:

$$\text{bias} = \frac{1}{N} \sum (\hat{V} - V) \quad (4.88)$$

$$\text{bias\%} = 100 \times \frac{\sum(\hat{V} - V)}{\sum V} \quad (4.89)$$

$$\text{mad} = \frac{1}{N} \sum |\hat{V} - V| \quad (4.90)$$

$$\text{mad\%} = 100 \times \frac{1}{N} \sum \left| \frac{\hat{V} - V}{V} \right| \quad (4.91)$$

$$\text{rmse} = \left[\frac{1}{N} \sum (V - \hat{V})^2 \right]^{0.5} \quad (4.92)$$

$$\text{rmse\%} = 100 \times \frac{[N \sum (V - \hat{V})^2]^{0.5}}{\sum V} \quad (4.93)$$

where

bias\% = relative bias

mad = mean absolute deviation

mad\% = relative mean absolute deviation

rmse = root mean squared error

rmse\% = relative root mean squared error

V = observed volume

\hat{V} = predicted volume

N = number of predicted volumes in the given class

4.3 Results

The estimated BLUP of θ is a weighted average between the first- and the second-stage model and is referred to as the weighted predictor. The overall performances of models from both stages individually and the weighted predictor as estimators of the observed growth rates and of the future growth rates are presented. Even if the predictions of future growth rates were the main interest, knowing the model behaviour for estimating the observed growth rates was also of interest.

Estimation of the observed growth rates were compared for the three “estimators” (the first-stage model, the second-stage model, and the weighted predictor). Points of comparison included estimation of the mean value, prediction of the minimum and maximum growth rates and estimation of the standard deviation of the growth rates.

Table 4.4: Estimation of Observed Growth Rates

	N	Mean	Minimum	Maximum	Std. Dev.
observed rates	162	0.03732	-0.08542	0.27720	0.04470
first-stage estimator	162	0.03962	-0.04393	0.23615	0.04441
second-stage estimator	162	0.04370	0.00084	0.21485	0.03547
weighted predictor	162	0.04246	-0.02738	0.23047	0.04347

As expected, the first-stage estimation of the mean growth rate was not equal to the mean observed growth rate since the unknown variance structure of the error term was estimated (Table 4.4). The observed growth rates were overestimated on average by about 6%. Since the predicted rates were an average of the observed rates on a plot, it is normal that the extreme predicted values be closer to the mean rate. However, the first-stage estimates were as variable as the observed rates.

The second-stage estimator was even more biased. It is a FGLS estimator of the first-stage estimates. FGLS estimators are biased in finite samples which explains the difference between first-stage and second-stage estimates. The mean second-stage estimate is 10% larger than the first-stage estimate. When compared to the observed rates, the difference jumped to 17%. The second-stage estimator did not predict negative growth rates or extreme growth rates which is intuitively reasonable. Estimates were 20.6% less variable than the observed growth rates.

When the weighted average of the first-stage and the second-stage estimator were used to predict the observed growth rates, the mean growth rate was overestimated by 13.8%. The weighted predictor tended to push the individual estimator toward the mean. Thus, the minimum and maximum growth rates predicted with this predictor were closer to the average prediction than the observed minimum and maximum. The mean, minimum and maximum estimates from the weighted predictor were between the respective estimates from the first-stage and second-stage models. The predicted estimates were slightly less variable than the observed growth rates.

The three estimators were also compared on the basis of their respective efficiency to predict future volumes. The overall results showed that the weighted predictor performed better than the two other estimators (Table 4.5). Assuming a constant growth rate on a plot for a short period of time was a convenient approximation in the first-stage model. However, to assume that this estimate will stay constant in the future will tend to overestimate future growth rates. The overall results of the first-stage estimator showed this overestimation. The first-stage model also gave highly variable estimates. The relative root mean squared error was 30% (for an average volume of 555.4 m³/ha). The predicted volume were off by 20% on average.

In the second-stage model, the choice of the initial volume was arbitrary and caused some uncertainty in the model. This was reflected in the results. The second-stage estimator underestimated the actual volumes by 3.2%. This suggests the choice of the initial volume to fit the second-stage model was probably underestimating the true initial volume. The root mean squared error was much smaller than with the first-stage model. The mean absolute deviation followed the same trend and was to 7.1%.

With the weighted predictor, the bias was reduced to 1.1%. The variance was smaller with the estimate of the second-stage model than with the estimate of the first-stage model, which pulled the weighted predictions closer to the second-stage estimates than

Table 4.5: Overall Results from the First- and the Second-Stage Estimators and from the Weighted Predictor

Predictor	N	bias	bias%	rmse	rmse%	mad	mad%
		(m ³ /ha)		(m ³ /ha)		(m ³ /ha)	
First-stage Estimator	189	55.7	10.0	166.6	30.0	83.2	20.1
Second-stage Estimator	189	-17.6	-3.2	71.9	12.9	42.5	7.1
Weighted Predictor	189	-6.0	-1.1	51.4	9.3	35.9	7.5

to the first-stage estimates. The weighted predictor was both more precise and more accurate than the second-stage model. The root mean squared error was reduced to about 9% and the mean absolute deviation was slightly more than 7%.

Further details are only presented for the weighted predictor as this predictor performed better than the two other alternatives.

Except for the youngest plots, the model consistently underestimated the actual volumes. The overestimation for the young plots is caused by the first-stage estimate, which assumed that the growth rate on a plot is constant. The growth rate is rapidly decreasing in a young plot and projecting the observed growth rate will cause a strong bias. The growth rate levels off as the plot gets older, which explains why the bias is much smaller on plots that had their first observations at 20 years of age or later. As a general trend, it can be noticed that the bias increased with projection time. (Table 4.6). This trend of increasing bias is not as clear when the total lag is taken into consideration (Table 4.7). Projecting a plot more than 25 years after the first measurement yielded seriously underestimated volumes for plots over 30 years of age.

The relative variability of the predictions was more important for the young plots. For plots with an initial age less than 20 years, the root mean squared error reached about 20% for projections over 6 years. On the other hand, the variation of the predictions for plots with initial age over 40 years were always less than 10%, generally less than 5% (Table 4.8). The mean absolute deviations were fairly small for plots whose initial age were greater than 20 (Table 4.10). When the total lag was considered, the predictions were more variable as the total lag increased, usually greater than 10% when the total lag was longer than 26 years. The mean absolute deviations followed the same trends as the root mean square errors.

Table 4.6: Bias and (Relative Bias) by Initial Age and Projection Time Class

Projection Time (yrs)	Initial Age (yrs)							
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90
0-5	21.8 (9.4)	4.3 (1.3)	-27.2 (-5.6)	8.6 (1.3)	-7.2 (-1.5)	-11.4 (-2.1)	-15.0 (-1.9)	-5.1 (-0.6)
6-10	48.5 (15.6)	-1.2 (-0.3)	-45.8 (-9.9)	6.4 (0.9)	-12.7 (-2.4)	-24.4 (-4.2)	-28.1 (-3.3)	0.5 (0.1)
11+	92.7 (22.1)	-4.3 (-0.9)	-44.1 (-7.2)	5.4 (0.7)	-30.5 (-5.2)	-33.0 (-5.2)	-38.3 (-4.1)	37.8 (4.1)

Table 4.7: Bias and (Relative Bias) by Initial Age and Total Lag Class

Total Lag (yrs)	Initial Age (yrs)							
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90
10-15	29.7 (11.3)	6.8 (2.9)	-1.4 (-0.6)	8.6 (1.3)	-7.2 (-1.5)	-9.7 (-1.9)	-17.5 (-3.5)	-5.1 (-0.6)
16-20	66.4 (18.9)	0.7 (0.2)	-3.5 (-1.3)	6.4 (0.9)	-12.7 (-2.4)	-22.7 (-3.9)	-21.7 (-2.6)	0.5 (0.1)
21-25	163.5 (30.7)	-6.8 (-1.4)	-36.1 (-7.3)	5.4 (0.7)	-30.5 (-5.2)	-27.8 (-4.4)	-38.0 (-4.3)	37.8 (4.1)
26+	- -	1.4 (0.3)	-48.7 (-7.4)	- -	- -	-61.4 (-8.1)	-25.7 (-2.1)	- -

Table 4.8: Root Mean Squared Error and (Relative Root Mean Squared Error) by Initial Age and Projection Time Class

Projection Time (yrs)	Initial Age (yrs)							
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90
0-5	25.3 (11.0)	23.0 (6.8)	36.5 (7.5)	14.6 (2.2)	8.7 (1.8)	22.8 (4.3)	20.9 (2.6)	12.9 (1.5)
6-10	61.3 (19.7)	41.6 (10.4)	60.3 (13.0)	38.2 (5.2)	16.6 (3.2)	38.5 (6.6)	37.1 (4.3)	9.9 (1.1)
11+	123.1 (29.4)	68.7 (13.9)	81.0 (13.2)	51.5 (6.4)	36.3 (6.2)	60.6 (9.5)	53.3 (5.8)	65.6 (7.1)

Table 4.9: Root Mean Squared Error and (Relative Root Mean Squared Error) by Initial Age and Total Lag Class

Total Lag (yrs)	Initial Age (yrs)							
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90
10-15	40.3 (15.3)	8.7 (3.6)	1.4 (0.6)	14.6 (2.2)	8.7 (1.8)	20.8 (4.1)	24.2 (4.8)	12.9 (1.5)
16-20	93.7 (26.6)	21.5 (5.4)	3.5 (1.3)	38.2 (5.2)	16.6 (3.2)	36.5 (6.3)	26.0 (3.2)	9.9 (1.1)
21-25	164.2 (30.8)	56.2 (11.7)	47.8 (9.7)	51.5 (6.4)	36.3 (6.2)	45.7 (7.3)	50.2 (5.7)	65.6 (7.1)
26+	- -	66.8 (14.7)	87.2 (13.3)	- -	- -	106.5 (14.1)	48.2 (3.9)	- -

Table 4.10: Mean Absolute Deviation and (Relative Mean Absolute Deviation) by Initial Age and Projection Time Class

Projection Time (yrs)	Initial Age (yrs)							
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90
0-5	21.8 (10.2)	15.3 (4.8)	28.7 (6.1)	9.7 (2.2)	7.2 (1.4)	16.7 (3.2)	17.3 (2.6)	11.9 (2.0)
6-10	48.8 (15.1)	28.8 (7.0)	45.8 (10.3)	28.5 (5.7)	12.7 (2.6)	34.3 (6.0)	33.8 (4.5)	9.9 (1.3)
11+	97.9 (20.6)	51.1 (11.1)	67.1 (10.8)	40.6 (7.5)	30.5 (4.9)	47.5 (7.2)	44.8 (5.7)	53.6 (5.0)

Table 4.11: Mean Absolute Deviation and (Relative Mean Absolute Deviation) by Initial Age and Total Lag Class

Total Lag (yrs)	Initial Age (yrs)							
	0-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90
10-15	29.9 (10.8)	6.8 (3.9)	1.4 (0.6)	9.7 (2.2)	7.2 (1.4)	15.3 (3.0)	19.6 (3.9)	11.9 (2.0)
16-20	70.5 (19.2)	14.7 (3.4)	3.5 (1.3)	28.5 (5.7)	12.7 (2.6)	33.2 (5.9)	23.0 (3.6)	9.9 (1.3)
21-25	163.5 (30.7)	42.5 (8.6)	37.1 (8.1)	40.6 (7.5)	30.5 (4.9)	38.1 (6.0)	43.8 (5.9)	53.6 (5.0)
26+	- -	49.5 (12.3)	75.5 (11.7)	- -	- -	87.1 (12.3)	38.8 (3.0)	- -

Chapter 5

Discussion

5.1 The Basic Concept

Obtaining an intuitive approximation of future volumes on permanent sample plots is not very difficult when the projection time is short. When a few volume observations from a plot are available, the most intuitive technique for a non-forester would be to figure out an average current annual increment from the past periodic increments, multiply this estimate by the number of years, and add this number up the current volume.

Another technique is the "interest rate" technique; that is, instead of working with the absolute value of growth, it is also possible to work with the relative growth. The average interest rate is estimated from past measurements, and the current volume is compounded for the required number of years.

These techniques have some drawbacks, though. The current annual increment and the growth rate are not really constant over time. For instance, projecting any estimated past growth rate will overestimate future volumes. To make up for the drop in future growth rate, some factor must be introduced to reduce the estimated growth rate before compounding the current volume. This compensatory factor should be correlated with time as the decrease in growth rate is more important in young plots than in old ones.

An alternative estimate is to use the information given by the relationship between observed growth rates and some plot attributes. This relationship must be estimated from a collection of individuals. This technique, called regression, has become common with

the proliferation of statistical software. The growth rate of an individual is assumed to be the expected growth rate of all individuals with similar plot attributes. The difficulty with this technique is that the growth rate estimation is not plot specific but is based on some expected value from a collection of individuals.

The regression technique is theoretically unbiased but rather imprecise when applied to individual plots as it does not take into account the specific performance of the individual plot compared to the average plot. This is of little importance if it is assumed that the individual relative performance in the past is unrelated to the relative performance in the near future. However, if it is assumed that the past individual relative performance is a good indicator of future individual performances, then the regression technique omits useful information: the individual plot performance.

The weighted predictor combines the information that can be drawn from the regression technique with the average growth technique. In a sense, the second-stage model is the correction factor for the projection of the average growth rate, obtained from the first stage. It makes intuitive sense that future growth rates be highly correlated with both previous growth rates and future plot attributes. The weighted predictor recognizes this fact by using both types of information.

The two-stage model and the resulting weighted predictor have some limitations, however. Knowing these limitations is essential to avoid misuse or misunderstanding.

Theoretically, the weighted predictor will always be biased because of the approximation in the first-stage model. This absolute bias should decrease as the plot gets older since the growth rate levels off with time.

When few measurements are available on each individual, it might be dangerous to try to estimate the individual component of the weighted predictor. Past individual performances based on 2 or 3 observations might not be a good indicator of future individual

performances. Harville (1976) did not suggest a minimum number of observations to estimate the BLUP of a parameter. Therefore, there is no guarantee that short time-series are a valuable source of information for individual performances.

Disturbed plots were not included in this study. This is a limitation of the second-stage model. At this stage, with our current knowledge, it is difficult to explain catastrophic disturbances and even more difficult to predict them; the decision was made not to include the disturbed sample plots in the study. Therefore, the capability of the model to predict future growth rates in a disturbed environment was not tested.

5.2 The Statistical Model

A statistical model is a simplification of reality. It represents the experimenter's imperfect understanding of the system. Furthermore, this understanding is limited in its expression by mathematical tractability. A model must be identifiable to be estimated. This becomes a serious constraint when only minimal information is available to fit the model. Assumptions have to be made; these assumptions can be seen as a tractable expression of the modeler's partial knowledge about the system. Thus, by definition, a statistical model is a compromise between reality and tractability.

The first-stage model represented this kind of compromise. First, for each plot one observation was lost by using growth rate instead of volume, diminishing an already small number of observations. That brought the number of observations down to two for 42 plots out of 68. These two observations were covering ten years on average since the mean period interval was five years. The constant growth rate assumption was made because obtaining repeated measurements of the same parameter is essential for parameter estimation. It was considered reasonable since for most plots, ten years is still a short period of time. Of course, had all plots had at least 4 volume measurements (3

growth observations), a first-stage model predicting a decreasing growth rate over time could have been proposed, and better results probably be achieved. The assumption of constant growth rate was not a constraint due to the technique used but due to the nature of the data set.

Defining the structure of the random term in the first-stage model (and in the second-stage model, for that matter) was certainly a major problem. Very little attention has been paid in forestry to error structure in model-building. Due to the absence of previous work, models from other fields had to be investigated. Most structures suggested in the literature on time-series require an extensive data set to be estimated with a reasonable confidence. North American foresters usually do not have access to such data sets. Estimating short time-series is a difficult task. Lack-of-fit and serial correlation are not easily differentiated when individuals have been remeasured only a few times. Common-sense assumptions must be made to overcome this limitation.

The assumptions about the error terms in the first-stage model reflect this absence of knowledge about the error structure. Two overlapping periods should intuitively be correlated. It is less certain, though, that two nonoverlapping periods can be automatically considered independent. For instance, two periods that had similar extreme climatic conditions could well be correlated. Nevertheless, the model neglects this information. Another consequence of the simplification of the covariance matrix of the time-dependent error terms is the independence between consecutive growth observations on an individual. One can probably argue that the temporal or the unexplained error terms might be correlated between periods. Considering this correlation as negligible is probably oversimplifying. However, for short-time series, it is reasonable to neglect some information rather than to try to extract it by making dubious assumptions.

The second-stage model includes a biologically-reasonable estimate within the weighted predictor. Von Bertalanffy's curve, or its more general form, the Chapman-Richards

curve, is probably the most used growth model in biological sciences to model growth. The yield model, given by the integration of the growth model can be written:

$$V = b_1(1 - e^{b_2(t-b_3)})^3$$

where t is time.

The parameter b_1 is the asymptotic volume of the plot. All curves with common asymptotic volumes will move to this volume at a pace depending on the parameter b_2 . The parameter b_3 is a location parameter. Growth models based on Von Bertalanffy's curve suggest some trends toward full-stocking since an asymptotic volume is assumed. This trend can be noticed in the second-stage model. Two plots with identical site indices but different volumes will have different growth rates; the plot with the smaller volume will increase its volume at a faster rate than the other plot, with both plots ending up with the same volume.

The growth rate is assumed to be a function of two variables: site index and standing volume. Site index is a function of age and height, so these variables can be considered as being indirectly included in the model.

Neither site index nor initial volume are ideal variables to be included in a model. While site index is generally accepted as a good index of site quality, its estimation on young plots might not be the best method for assessing the potential of a site. The definition of standing volume is also subject to some variation. The definition used offered satisfactory results. With a similar problem, Hui and Berger (1983) used for each individual the point where a constant growth rate and a linear growth rate fitted individually would have given an identical estimate. This technique was tried but it sometimes yielded volume estimates that were outside the range of observed volumes on a plot, even giving a negative estimate for one plot. These problems ruled out this definition. The chosen definition for initial volume was reasonable and sufficiently efficient.

Age, a common variable in growth and yield modeling, was not directly included in the second-stage model. Growth was assumed to be a result of photosynthetic activity. This activity is executed by cells that are never more than a few year old at any point in time in the life of a plot. Growth is undoubtedly correlated with age, but it is not caused by age. Age is simply correlated with the real cause of growth, the standing volume.

Density, another variable commonly used to predict growth, is not included in the model either. The most common forms of expressing density are probably the number of trees per hectare, basal area, and percentage of crown closure. It was believed that the volume per hectare could compensate for this variable since it can also be considered as an expression of density.

When the second-stage model is combined with the first-stage model, individual growth rates can be projected into the future.

While the error structure of the combined model was believed to be satisfactory for short projections, it might be inadequate for long projections. In the weighted predictor, the inverse of the variance of an estimate from one stage is a measure of the confidence in this stage's estimate. As the projection lag increases, the modeler's confidence in the first-stage model should diminish; the first-stage estimate should be closer to the actual growth rate in the first projection year than in the last one. Therefore, the variance of the estimate should reflect this change in confidence. In the weighted predictor, the relative confidence of the first-stage estimate (compared to the second-stage estimate) decreases as the confidence in the second-stage estimate improves with increasing volume, but it does not change in absolute value. However, this oddity was not important enough on short time-series to be corrected. If longer projection lags had been available, this problem would have been addressed.

The purpose of the weighted predictor was to find a reasonable compromise between accuracy, precision, and mathematical tractability for projecting permanent sample plots

when minimal information is available to the modeler. The model was not intended to predict long-term yield or predict yield from an extensive data set. The model uses as much information as can be used and gives limited but useful information. Fortunately, the concept of a two-stage framework can be extended easily to plots with more measurements. Another advantage of the two-stage framework is that the second stage constitutes the best option to project temporary sample plots, or permanent sample plots with less than three measurements.

The combined model can be considered a random coefficients model. A random coefficients model recognizes that observations from an individual are random variations around the true individual relationship and that individual relationships are random variations around a general model. Permanent sample plots allow modeling these random relationships. The difference between random observations and random relationships has rarely been made in forestry. Growth is not an instantaneous variable such as volume or basal area, for instance. It is a linear variable, and it must be estimated as a line.

Geometrically speaking, instantaneous variables are dimensionless, while linear variables have one dimension. Regression through a series of dimensionless points does not yield a line but an infinite series of points. The expected value of a dimensionless variable is a dimensionless value. This can easily be illustrated by an example of a simple linear regression where the only possible x-values are discrete. Then, if the scale of the x-axis is large enough, there will be gaps between predicted points since prediction for values in between discrete x-values are meaningless. The scale of the x-axis can also be reduced in such a way that the predicted points form a line if the range of x-values is large enough. However, regression through a series of lines does yield a line. The expected value of a one-dimensional variable is a one-dimensional value. This is a major difference when dealing with permanent sample plots.

Recognizing the random nature of the relationship between observations on a plot, and that this relationship is the parameter of interest, are essential steps in modeling with permanent sample plots. This motivates the use of a random coefficient model. Unfortunately, the random coefficient model has not been studied as thoroughly as the components of variance model or the seemingly unrelated regressions model due to the extra complexity of the model (Dielman 1989, p. 95), but there has been a renewed interest (Gumpertz and Pantula 1989). Meanwhile, the lack of theoretical knowledge can be partially compensated by simulation or empirical evidence. Hopefully, more frequent use of random coefficient models will stimulate more research to improve our understanding of these models.

5.3 Results

Fitting the first-stage model was relatively fast and required a little more than 1 Mb of memory. The two design matrices (\mathbf{F} and \mathbf{G}) with the covariance matrix of the estimated rates ($[\mathbf{F}'\mathbf{\Omega}^{-1}\mathbf{F}]^{-1}$) used most of the memory. Results from the first-stage estimators were highly predictable. It was the best estimator of the observed growth rates but the poorest for predicting future rates. This estimator could adequately report what happened in the past, but any extrapolation to the future would be dangerous. This is best illustrated with negative growth rates. The first-stage estimator will yield negative estimates for plots undergoing a reduction in volume between the first and the last measurement. However, these negative growth rates can be considered short-term adjustments, and are not an accurate indicator of future growth rate. The first-stage estimate is merely an observation rather than an explanation of the growth rate.

Convergence of the second-stage model was also fast; only six iterations were needed before convergence. The initial value used for ϕ^2 was 0. The second-stage was the poorest

estimator of observed rates. To make comparison possible with the first-stage model, the second-stage model was used to predict $\hat{\theta}_i$, not the individual observations. Better results could have probably been achieved by using the observed initial volumes and by trying to predict the observed rates. Another reason for the weak results in the second stage is that it could not predict extreme growth rates. This poor performance was also expected since the number of plots was not large and the estimators of β and, especially, ϕ^2 were based on large sample theory. Surprisingly, the second-stage model could predict future growth rates with respectable precision but a rather large bias. Overall, the second-stage model underestimated the volume, suggesting that the initial volume used to fit the model was possibly underestimated.

Writing the code for the prediction of future volumes was simple using a matrix algebra language. When no variance estimates are computed, the predictions are rapidly obtained. However, computing the confidence intervals of the predictions can be cumbersome. The overall results of the weighted predictor were always between the overall results from the first-stage and the second-stage models. The weighted predictor also did poorly in estimating the observed mean growth rates. This is not surprising since it is derived from the best estimate for observed rates, the first-stage estimate, to which a certain quantity was added. Combining the first- and the second-stage was of no use for fitting the observed growth rates. If fitting past observations had been the only purpose, the first-stage estimator would have been sufficient.

When future volumes were considered, the weighted predictor was more accurate than either of the individual estimates. The variation of the prediction errors was considerably better than the first-stage estimate and slightly smaller than the second-stage model. The weighted predictor was by far the best option for predicting future growth rates.

Analyzing the detailed results, it is seen that the relative bias was more important for predictions on plots less than 20 years of age initially. For projection time less than

10 years, on plots with initial ages less than 20 years, the bias was close to the negative value of the mean absolute deviation, suggesting that the weighted predictor consistently overestimated the actual volumes. However, because these plots have less volume, they carry only a small weight when the overall bias is computed. For plots with initial age between 31 and 40 years, when only the projection time was analyzed, there was a strong bias. This probably occurred by chance, no rationale could be found to explain such a behaviour. Curiously, the length of the total lag did not seem to have an effect on the accuracy of the predictions. The projection time had a stronger influence on accuracy.

The precision of the estimates was dependent on both the projection time and the total lag. There was a major difference in precision between plots with initial age less than 40 years and plots with initial age over 40 years. This probably indicated where the curve of growth rate over time for Douglas-fir in the data set began to level off. For plots with initial age over 40, the variability was less than 10% most of the time. For these plots, the first-stage estimate yielded precise estimates which explains the good precision of the weighted predictor.

The mean absolute deviation is the arithmetic average of the error terms in absolute value. For plots over 40 years of initial age, it was usually less than 5%, even for projection times greater than 11 years. It was also relatively larger on young plots. This lack of precision was mainly caused by the first-stage estimates, as confirmed by looking at the results in detail. The precision of the second-stage estimator was relatively constant.

Chapter 6

Conclusions

Statistical models based on a collection of remeasured entities are common in a few scientific fields. However, permanent sample plots in forestry have particular complexities that models developed in other fields do not address. Adequate techniques must be developed to predict growth in forestry permanent sample plots. These techniques must be adaptable to plots with only a few measurements since this is the situation most often encountered in Canada. Above all, these techniques must rely on a clear understanding of the purposes of permanent sample plots, and their possibilities and limitations as sources of information.

The framework presented in this dissertation reflects this understanding. Growth rates were explained by site index and initial volume in a model inspired from Von Bertalanffy's growth curve. The individual relationship characterizing a sample plot was expressed as a random variation around a general model. Observations on a plot were assumed to be random outcomes of the plot's individual model. The current volume of a plot was updated using the growth rate predicted from the individual model. This technique was shown to be more precise than using the general model common to all plots.

The weighted predictor also emphasizes the importance of the error structure in forestry statistical models, as pointed out by Gregoire (1987). Forestry information is not equally valuable; this fact must be recognized in a model. Separating the error term into components facilitates assigning a specific weight to each piece of information.

A greater weight must be given to reliable information when computing an estimator. Precision of an estimator is as important as its bias. However, too often foresters have not paid attention to precision. Accurate mathematical models are sought to the detriment of statistical models. The power of statistics lies in its recognition of the difference between a variable and a constant. Statistical models will always be more precise than mathematical models when the error structure is correctly specified. Because models are used to make decisions, precise models are needed to make well-informed decisions.

The results of the numeric example illustrated the advantages of the weighted predictor. Despite the approximations made in both stages, the weighted predictor was reasonably accurate and precise. It is probable that using a larger number of plots would have resulted in more precision with the second-stage model and consequently to the weighted predictor as well. Permanent sample plots with a longer history would also have provided more precise results. Fortunately, the weighted predictor can be adapted to data sets where all sample plots have more than three measurements.

The weighted predictor is certainly not perfect. A major problem with the weighted predictor is the absence of statistical theory dealing with random coefficient modeling in finite samples. The estimator is approximately precise, but how approximate is unknown. The first-stage approximation can be justified for a short period of time only. No stochastic term is included for predicting catastrophic losses. Despite these weaknesses, the weighted predictor can provide useful information for solving problems facing many forest agencies in Canada.

The proposed framework, which led to the weighted predictor, opens new horizons for modeling using permanent sample plots. Developments in the theory related to random coefficient models in finite samples can be expected in the near future and should answer many important questions. New computer technology will facilitate large simulation problems. Research on techniques to include probabilities of catastrophic losses in

prediction could improve the decision-making process based on growth models. Foresters will benefit from these improvements if they adopt the right framework. Perfect models might still be far ahead, but we are definitely getting closer.

References Cited

- Aitchison, J. and Brown, J. A. C. 1957. *The lognormal distribution*. Cambridge Univ. Press, Cambridge, England, 176 p.
- Aitken, A. C. 1935. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh* 55:42-48.
- Amemiya, T. 1978. A note on random coefficient model. *Inter. Econom. Rev.* 19:793-796.
- 1971. The estimation of the variances in a variance-component model. *Internat. Econom. Rev.* 12:1-13.
- Arner, S. L. and Seegrist D. W. 1980. Missing: a computer program for the maximum likelihood estimates of the parameters of the multivariate linear model with incomplete measurements. *USDA For. Serv. Gen. Tech. Pap. NE-56*, 19 p.
- 1979. A computer program for the maximum likelihood estimator of the general multivariate linear model with correlated errors. *USDA For. Serv. Gen. Tech. Pap. NE-51*, 10 p.
- Arora, S. S. 1973. Error components regression models and their applications. *Ann. Econom. Social Measurement* 2:451-461.

- Axelsson, E. and Axelsson, B. 1986. Changes in carbon allocation patterns in spruce and pine trees following irrigation and fertilization. *Tree Physiology* 2:189-204.
- Baltagi, B. H., 1981. Pooling: an experimental study of alternative testing and estimation procedures in a two-way error component model. *J. Econometrics* 17:21-49.
- Baskerville, G. L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* 2:49-53.
- Bayes, T. R. 1763. An essay towards solving a problem in the doctrine of chances. *Philo. Trans. Roy. Soc.* 53:370-418 (reprinted in *Biometrika* 45:296-315).
- Berkey, C. S. 1982. Bayesian approach for a nonlinear growth model. *Biometrics* 38:953-961.
- Biging, G. S. 1985. Improved estimates of site index curves using a varying-parameter model. *For. Sci.* 31:248-259.
- Buckman, R. R. 1962. Growth and yield of red pine in Minnesota. *USDA For. Serv. Tech. Bull.* 1272, 50 p.
- Carter, R. L. and Yang, M. C. K. 1986. Large sample inference in random coefficient regression models. *Comm. Statist. A:Theory-Methods* 15:2507-2525.
- Chapman, D. G. 1961. Statistical problems in population dynamics, *In Proc. Fourth Berkeley Symp. Math. Stat. and Prob.*, Univ. Calif. Press, Berkeley, p. 153-168.

- Clutter, J. L. 1963. Compatible growth and yield models for loblolly pine. *For. Sci.* 9:354-371.
- Curtis, R. O. 1967. A method of estimation of gross yield of Douglas-fir *For. Sci. Monogr.* 13, 24 p.
- Davis, L. S. and Johnson, K. N. 1987. *Forest management*. Third edition. MacGraw-Hill Book Company, New York, 790 p.
- Davis, A. W. and West, P. W. 1981. Remarks on "Generalized least squares estimation of yield functions" by I. S. Ferguson and J. W. Leech. *For. Sci.* 27:233-239.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm.(with discussion) *Jour. Roy. Stat. Soc. series B* 39:1-38.
- Dhrymes, P. J. 1971. Equivalence of iterative Aitken and maximum likelihood estimators for a system of regression equations. *Austral. Econom. Papers* 10:20-24.
- Dielman, T. E. 1989. *Pooled cross-sectional and time-series data*. Marcel Dekker Inc., New York, 249 p.
- 1983. Pooled cross-sectional and time-series data: a survey of current statistical methodology. *Am. Stat.* 37:111-122.
- Fearn, T. 1975. A Bayesian approach to growth curves. *Biometrika* 62:89-100.
- Ferguson, I. S. and Leech, J. W. 1978. Generalized least squares estimation of yield functions. *For. Sci.* 24:27-42.

- _____. 1981. Reply to remarks by A. W. Davis and P. W. West on "Generalized least squares of yield functions." *For. Sci.* 27:589-591.
- Fuller, W. A. and Battese, G. E. 1974. Estimation of linear models with crossed-error structure. *J. Econometrics.* 2:67-78.
- Furnival, G. M. and Wilson, R. W. Jr. 1971. Systems of equations for predicting forest growth and yield. *In* Statistical Ecol. 3:43-57. G. P. Patil, E. C. Pielou, W. E. Waters, eds. Penn State Univ. Press, University Park.
- Garcia, O. 1983. A stochastic differential equation model for the height growth of forest stands. *Biometrics* 39:1059-1072.
- Geisser, S. 1970. Bayesian analysis of growth curves. *Sankhyā A* 32:53-64.
- Gerhardt, E. 1930. *Ertragstafeln für reine und gleichartige Hochwaldbestände von Eiche, Buche, Tanne, Fichte, Kiefer, grüner Douglasie und Lärche*. Second edition. Springer-Verlag, Berlin.
- Goldberger, A. S. 1968. The interpretation and estimation of Cobb-Douglas functions. *Econometrica* 36:464-472.
- Gregoire, T. G. 1987. Generalized error structure for forestry yield models. *For. Sci.* 33:423-444.
- Grizzle, J. E. and Allen, D. M. 1969. Analysis of growth and dose response curves. *Biometrics* 25:357-381.

- Gumpertz M. and Pantula, S. G. 1989. A simple approach to inference in random coefficient models. *Am. Stat.* 43:211-215.
- Harville, D. 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Stat.* 4:384-395.
- Hui, S. L. and Berger, J. O. 1983. Empirical Bayes estimation of rates in longitudinal studies. *J. Am. Stat. Ass.* 78:753-760.
- Judge, G. G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., and Lee, T. C. 1985. *The theory and practice of econometrics*. Second edition. John Wiley and Sons, New York, 1019 p.
- Keyes, M. R. and Grier, C. C. 1981. Above- and below-ground net production in 40-year old Douglas-fir stands on low and high productivity sites. *Can. J. For. Res.* 11:599-605.
- King, J. E. 1966. Site index curves for Douglas-fir in the Pacific Northwest. Weyerhaeuser Co. For. Res. Center, For. Paper 8. 50 p.
- Kmenta, J. and Gilbert R. F. 1970. Estimation of seemingly unrelated regressions with autoregressive disturbances. *J. Am. Stat. Ass.* 65:186-197.
- Kurz, W. A. and Kimmins, J. P. 1987. The influence of site quality on tree resource allocation to fine roots and its effect on harvestable productivity of coastal Douglas-fir stands. Can. For. Service, Victoria, B.C. FRDA report 034, 102 p. (+ appendices)

- Laird, N. M. and Ware, J. H. 1982. Random-effects models for longitudinal data. *Biometrics* 38:963-974.
- Leak W. 1966. Analysis of multiple systematic remeasurements. *For. Sci.* 12:69-73.
- Lee, L. F. and Griffiths, W. E. 1979. The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models. Univ. of New England Working Papers in Econometrics and Applied Statistics No. 1, Armidale, Australia.
- LeMay, V. 1988. Comparison of fitting techniques for systems of forestry equations. Unpublished Ph.D. dissertation, Univ. of B. C., 172 p.
- Lindley, D. V. and Smith, A. F. M. 1972. Bayes estimates for the linear model. *J. Roy. Stat. Soc. B* 34:1-41.
- Maddala, G. S. 1977. *Econometrics*. McGraw-Hill, New York, 516 p.
- 1971. The use of variance components models in pooling cross-section and time-series data. *Econometrica* 39:341-358.
- Maddala, G. S. and Mount, T. D. 1973. A comparative study of alternative estimators for variance component models used in econometric applications. *J. Am. Stat. Ass.* 68:324-328.
- Marshall, P. L. and Jahraus, K. V. 1987. Growth and yield availability and use in British Columbia. Unpublished report, 96 p.

- Mundlak, Y. 1978. Models with variable coefficients: integration and extension. *Ann. de l'INSEE* 30-31:483-509.
- 1961. Empirical production functions free of management bias. *J. Farm Econom.* 43:44-56.
- Parks, R. W. 1967. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *J. Am. Stat. Ass.* 62:500-509.
- Phillips, P. C. B. 1985. The exact distribution of the SUR estimator. *Econometrica* 53:745-756.
- Pienaar, L. V. and Turnbull, K. J. 1973. The Chapman-Richards generalization of Van Bertalanffy's growth model for basal area growth and yield in even-aged stands. *For. Sci.* 19:2-22.
- Pothoff, R. F. and Roy, S. N. 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51:313-326.
- Rao, C. R. 1987. Prediction of future observations in growth curve models. *Stat. Sci.* 2:434-471.
- 1975. Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics* 31:545-554.
- 1972. Estimation of variance and covariance components in linear models. *J. Am. Stat. Ass.* 67:112-115.

- _____. 1965. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* 52:447-458.
- Rao, U. L. G. 1982. A note on the unbiasedness of Swamy's estimator for the random coefficient regression model. *J. Econometrics* 18:395-401.
- Richards, F. J. 1959. A flexible growth function for empirical uses. *J. Exp. Bot.* 10:290-300.
- SAS Institute Inc. 1982. *SAS user's guide: statistics, 1982 edition*. SAS Institute Inc., Cary, NC, 584 p.
- Schmidt, P. 1977. Estimation of seemingly unrelated regressions with unequal number of observations. *J. Econometrics* 5:365-377.
- Schumacher, F. 1939. A new growth curve and its applications to timber-yield studies. *J. For.* 37:819-820.
- Seegrist, D. W. and Arner, S. L. 1978. Statistical analysis of linear growth and yield models with correlated observations from permanent plots remeasured at fixed intervals. p 209-233 *In* Growth models for long-term forecasting of timber yields, (eds. J. Fries, H. Burkhart, T. A. Max), Sch. For. Wild. Res. VPISU, Publ. FWS 1-78.
- Smith, A. F. M. 1973. A general Bayesian linear model. *J. Roy. Stat. Soc. B* 35:67-75.
- Srivastava, V. K. 1973. The efficiency of an improved method of estimating seemingly unrelated regression equations. *J. Econometrics* 1:341-350.

- Sullivan A. D. and Clutter, J. L. 1972. A simultaneous growth and yield model for loblolly pine. *For. Sci.* 18:76-86.
- Sullivan A. D. and Reynolds M. R. Jr. 1976. Regression problems from repeated measurements. *For. Sci.* 22:382-385.
- Swamy, P. A. V. B. 1974. Linear models with random coefficients, *In* *Frontiers in econometrics* (ed. P. Zarembka), New York, Academic Press, p. 143-168.
- 1971. *Statistical inference in random coefficient regression models*. Springer-Verlag, New York, 209 p.
- 1970. Efficient inference in a random coefficient regression model. *Econometrica* 38:311-323.
- Swamy, P. A. V. B. and Arora, S. S. 1972. The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica* 40:261-275.
- Swamy P. A. V. B. and Metha J. S. 1979. Estimation of common coefficients in two regression equations. *J. Econometrics* 10:1-14.
- 1975. On Bayesian estimation of seemingly unrelated regressions when some observations are missing. *J. Econometrics* 3:157-169.
- 1973. Bayesian analysis of error components regression. *J. Am. Stat. Ass.* 68:648-658.

- Swindel, B. F. 1968. On the bias of some least squares estimators of variance in a general linear model. *Biometrika* 55:313-316.
- Tait, D. E. N., Cieszewski, C. J., and Bella, I. E. 1988. The stand dynamics of lodgepole pine. *Can. J. For. Res.* 18:1225-1260.
- Taylor, W. E. 1980. Small sample considerations in estimation from panel data. *J. Econometrics* 13:203-223.
- Ung, C. H. and Végiard, S. 1988. Problèmes d'inférence statistique reliés à la transformation logarithmique en régression. *Can. J. For. Res.* 18:733-738.
- Wallace, T. D. and Hussain, A. 1969. The use of error component models in combining cross-section with time-series data. *Econometrica* 37:55-72.
- Zellner, A. 1987. *An introduction to Bayesian inference in econometrics*. R. E. Krieger Publishing Co., Malabar, FL, 431 p.
- 1962. An efficient method for estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Stat. Ass.* 57:38-368.
- Zellner, A. and Vandaele, W. 1975. Bayes-Stein estimators for k-means, regressions, and simultaneous equation models *In Studies in Bayesian Econometrics and Statistics* (ed. S. E. Fienberg and A. Zellner), Amsterdam, North Holland, p. 627-653.