Estimating the Coefficients in a System of Compatible Growth and Yield Equations for Loblolly Pine.

by

Richard P. Hans

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Forestry

APPROVED:

Richard G. Oderwald, Chairman

Harold E. Burkhart

David Wm. Smith

August, 1986

Blacksburg, Virginia

Estimating the Coefficients in a System of Compatible Growth and Yield Equations for Loblolly Pine.

by

Richard P. Hans

Richard G. Oderwald, Chairman

Forestry

(ABSTRACT)

In this thesis the five equation system of growth and yield equations originally developed by Clutter(1963) is examined. The system is redeveloped algebraically to form a truly algebraically compatible system.

Three methods of estimating the coefficients were examined. In the first method, three of the equations were fitted independently using ordinary least squares; these coefficient estimates were carried through to the other equations. No consideration was given to the relationships that must exist between the equation coefficients in order for the system to be numerically consistent. In the second method the system is first developed algebraically, before any of the coefficients are estimated, resulting in a slightly different system which is truly algebraically compatible. The coefficients were estimated by fitting two of the equations, and using these estimates throughout the rest of the system. The resulting system is both numerically consistent and algebraically compatible. In the final method the relationships between the coefficients that must hold for the system to be compatible were incorporated in the coefficient estimation procedure. Seemingly unrelated regression techniques were used to estimate the coefficients.

The three methods resulted in coefficient estimates that were similar, with seemingly unrelated regression producing the most efficient estimators. Prediction ability of the three methods on independent data show no method as being superior, although the seemingly unrelated regression procedure was able to reduce the total system error best.

Acknowledgements

I would like to thank Dr. Richard G. Oderwald for his suggestions and guidance throughout this study. I also wish to thank Dr. Harold E. Burkhart, and Dr. David Wm. Smith for their input, and advise as committee members. I also would like to thank Dr. Timothy G. Gregoire for his suggestions and advice.

I also wish to thank Mr. Thomas A. Dierauf, and the Virginia Division of Forestry for supplying the data used in this study.

To the people of room 321, thanks for your friendship and encouragement through the past two years. Thanks to my parents for their encouragement throughout my educational years, and a very special thanks to my wife Mary, for without her love and support, this thesis would not exist.

Table of Contents

List of Illustrations

List of Tables

Chapter 1

Introduction

There have been numerous publications on growth and yield in the past. Earliest methods used graphical techniques to predict growth and yield, then, with the increasing availability of computers, regression became the most commonly used technique for estimating the coefficients in growth and yield equations. Regression today is still the most widely used method for estimating coefficients.

Since the middle fifties it has been common practice to select candidate models for growth and yield and use regression techniques to refine the model and estimate equation coefficients for a specific data set. Many improvements in model specification and development have been made, most notable in the algebraic and numerical compatibility of growth and yield systems. However, problems with compatibility in the regression method of coefficient estimation arise because of the least squares principle of regression optimizes coefficients only for the dependent variable under consideration. If coefficients are estimated for a volume growth equation and a volume yield equation based on the growth model, coefficient values in the two equations will differ because of the difference in the sum of squares being minimized. Therefore, while the two equations are algebraically compatible, they are not numerically compatible.

It may be important for a forest manager to have a system of equations that is not only algebraically compatible, but numerically consistent as well. That is, if basal area is predicted out to a future age and this age and basal area are then used in the volume yield equation, the resultant volume should be equal to the volume obtained from the volume prediction equation using the future age and initial conditions. In order for the system to be numerically consistent, the relationships between the coefficients from one equation to the next must be obtained, and when the coefficients in each equation are estimated, the coefficients must be constrained for the system to be consistent.

The system of equations looked at in this study can be estimated using seemingly unrelated regression. Seemingly unrelated regression will be compared to the more commonly used method of ordinary least squares to see if an improvement in the estimation of the coefficients can be made.

Objectives:

1.) Develop a system of growth and yield equations that are numerically as well as algebraically consistent.

2.) Estimate the coefficients of the system simultaneously, using seemingly unrelated regression.

3.) Compare the system of equations with the coefficients estimated using seemingly unrelated regression to a compatible and a non-compatible system estimated using ordinary least squares.

Chapter 2

Literature review

Clutter(1963) and Buckman(1962) where the first researchers to express the differential/integral relationship between growth and yield. Using this relationship Clutter(1963) developed a system of compatible analytic models for total stand cubic-foot growth and yield for lobolly pine. He defined compatible to be when the algebraic form of the yield model can be derived by mathematical integration of the growth model.

Clutter(1963) started with the following two equations.

Volume yield equation

[1] $\ln(V) = \beta_0 + \beta_1 S + \beta_2 \ln(B) + \beta_3 A^{-1}$

Basal Area Growth Rate

[2] $\frac{\partial B}{\partial A} = -B\,Ln(B)\,A^{-1} + c_0 B\,A^{-1} + c_1 B\,S\,A^{-1}$

where

B = basal area per acre

S = site index

A = age of the stand

V = total cu.ft. volume

Both of these equations were fitted using linear regression to find estimates of the coefficients. By differentiating the volume yield equation with respect to age, a growth rate equation for volume was developed.

$$\frac{\partial V}{\partial A} = -\beta_3 V A^{-2} + \beta_2 V B^{-1} \frac{\partial B}{\partial A}$$

Substituting the basal area growth rate equation for $\frac{\partial B}{\partial A}$ the growth rate equation becomes.

[3] $$\frac{\partial V}{\partial A} = -\beta_2 V Ln(B) A^{-1} + \beta_2 c_0 V A^{-1} + \beta_2 c_1 V S A^{-1} - \beta_3 V A^{-2}$$

This equation was also fitted using linear regression.

Integrating both the basal area and volume growth rate equations from initial age to a predicted age and from initial basal area/volume to predicted basal area/volume and solving for the predicted basal area/volume results in basal area and volume prediction equations.

By using regression to fit the equations independently at different stages in the development of the system the numerical consistency of the system fails. The system is assumed to be algebraically consistent, but by fitting the equations independently at different stages and using the fitted values for the equation coefficients as the system is developed, results in a system that is not only numerically inconsistent, but also algebraically inconsistent.

Sullivan and Clutter(1972) noted that the coefficients for Clutter's(1963) set of equations are not independent between the equations. By not taking the dependence into account when estimating the coefficients, the system of equations are not numerically consistent as was the case in Clutter's(1963) article. Also measurements of such variables as stand volumes and basal area on the same plot do not constitute statistically independent observations. The authors developed a method to estimate the coefficients which takes both these problems into account.

Starting with Clutter's(1963) cubic-foot yield and projected basal area equations

$$[4] \qquad \ln(V_1) = \beta_0 + \beta_1 S + \beta_2 A_1^{-1} + \beta_3 \ln(B_1)$$

$$[5] \qquad \ln(V_2) = \beta_0 + \beta_1 S + \beta_2 A_2^{-1} + \beta_3 \ln(B_2)$$

$$[6] \qquad \ln(B_2) = (\frac{A_1}{A_2}) \ln(B_1) + \alpha_1(1 - \frac{A_1}{A_2}) + \alpha_2(1 - \frac{A_1}{A_2})S$$

where

A_1 = initial age

A_2 = future age

B_1 = basal area at age A_1

B_2 = basal area at age A_2

V_1 = volume at age A_1

B_2 = volume at age A_2

and substituting equation[6] into equation[5] for the $\ln(B_2)$ term, a projected cubic-foot volume equation can be written as

$$[7] \qquad \ln(V_2) = \beta_0 + \beta_1 S + \beta_2 A_2^{-1} + \beta_3(\frac{A_1}{A_2})\ln(B_1) + \beta_4(1 - \frac{A_1}{A_2}) + \beta_5(1 - \frac{A_1}{A_2})S$$

When $A_1 = A_2$ the equation collapses to the yield equation, therefore the equation is simultaneously a yield model and projection model. The coefficients β_4 and β_5 must be equal to $\beta_3 \times \alpha_1$ and $\beta_3 \times \alpha_2$ respectively. When the coefficients are found with this relationship between the coefficients the equations will be numerically consistent. Sullivan and Clutter(1972) suggested fitting equation [7] to find estimates for the b's, and then use the relationship $\alpha_1 = \frac{\beta_4}{\beta_3}$ and $\alpha_2 = \frac{\beta_5}{\beta_3}$ to estimate the coefficients of the basal area growth rate equation[6].

When developing estimators for the coefficients of equation [7] certain assumptions about the samples from the first and second measurements must be made. Sullivan and Clutter(1972) made the following assumptions: the observations in the first measurement will have a common variance σ_1^2 ; the observations from the second measurement will have a common variance σ_2^2 ; the covariance between different plots is zero; and there is a common covariance p, between the first and second measurements on the same plot. Under these assumptions the least square estimate of β is neither the maximum likelihood estimator nor the best linear estimator. Therefore an alternative method of estimation was used. Using the principle of maximum likelihood and assuming a bivariate normal distribution, estimators were developed by the authors. Although the derivation of the maximum likelihood estimates for the coefficients is more involved then using linear regression, the assumptions made are more reasonable. In the case of Sullivan and Clutter(1972) this difference was probably not of practical significance, but in other cases with different data the use of the maximum likelihood procedure may be necessary.

Burkhart and Sprinz(1984) used an alternative method to estimate the coefficients in Clutter's(1963) equations. They used the equations written in the same form as Sullivan and Clutter(1972), equations 1-4, noting that $\alpha_1 = \frac{\beta_4}{\beta_3}$ and $\alpha_2 = \frac{\beta_5}{\beta_3}$ must be true for the system to be numerically consistent. Since these estimates of α_1 and α_2 are not the most statistically efficient estimators (none of the basal area growth data is included) and they are dependent on the volume units and merchantability limits chosen for the dependent variable, Burkhart and Sprinz(1984) suggested another way to estimate the coefficients. They fitted the basal area equation [6] and the growth rate equation [7] simultaneously. The criteria to be minimized in the fitting gave equal weights to volume and basal area and is defined as follows.

$$F = \Sigma \frac{(y_i - \hat{y}_i)^2}{\hat{\sigma}_Y^2} + \Sigma \frac{(B_i - \hat{B}_i)^2}{\hat{\sigma}_B^2}$$

Where y_i and B_i are the observed values for volume and basal area and $\hat{y}_i$ and $\hat{B}_i$ are the predicted values. $\hat{\sigma}_Y^2$ and $\hat{\sigma}_B^2$ are estimated using the mean square error from the least square fits of the volume and basal area equations, respectively. The restrictions $\alpha_1 = \frac{\beta_4}{\beta_3}$ and $\alpha_2 = \frac{\beta_5}{\beta_3}$ were imposed. This fitting procedure resulted in coefficients for the basal area equations that were almost identical regardless of merchantability definition used in the yield equation and equations that were numerically consistent.

Curtis(1967) also used the relationship of growth rate expressed as a differential equation and the yield equation equal to the integral of the growth rate. He started with the growth rate equation

$$\frac{\partial Y}{\partial A} = f(A)f(S)f(RD)$$

where A = age, S = site, RD = relative density, and Y = gross volume or basal area. A gross cubic foot volume yield equation was then obtained by integrating the growth rate equation. Comparisons between estimates of growth rate and yield on permanent plots to those obtained on temporary plots showed that satisfactory estimates can be obtained from the temporary plots.

The previous procedures for developing yield functions involved integrating the growth rate equation and expressing the yield equation as a function of stand age. Moser and Hall(1969) noted that stand age has no meaning in uneven-aged stands, which makes the previous methods difficult to apply. Therefore they proposed using a growth rate equation which is expressed as a function of a measurable size characteristic of the quantity under investigation. Using time, since age cannot be determined, the growth rate is expressed as

[8] $$\frac{\partial Y}{\partial T} = f(Y)$$

where Y=cumulative growth and T=elasped time. Volume was then expressed as a function of basal area.

[9] $$V = \beta_0 B^{\beta_1}$$

Differentiating this equation with respect to time results in the growth rate equation.

[10] $$\frac{\partial V}{\partial T} = \beta_1 V B^{-1} \frac{\partial B}{\partial T}$$

The basal area growth rate is based on Von Bertalanffy's generalized growth rate equation.

[11] $$\frac{\partial B}{\partial T} = nB^{m} - kB$$

Substituting $\frac{\partial B}{\partial T}$ into equation [10] the final growth rate equation that results is

$$[12] \quad \frac{\partial V}{\partial T} = \beta_1 V(nB^{m-1} - k)$$

The equations were then fitted using regression. The volume and basal area yield functions were then found by solving the respective growth rate equations for volume and basal area. This article provided appropriate methodology for deriving time dependent yield functions from growth rate equations which do not have time or age as an independent variable.

When developing a compatible system of equations, the best empirically derived growth rate equation, when integrated, may not result in the best yield equation. Smith(1983) derived an empirical basal area growth rate equation and two constrained growth rate equations based on Clutter's(1963) and Buckman's(1962) basal area growth rate equations. He found although the empirical growth rate equation fit the data best, when the compatible yield equations where examined, the yield equation derived from Buckman's(1962) constrained growth rate equation gave the best predictions of future yields. This result gives some justification for using the constrained model over an empirically derived one.

A different approach which relies on restricted three-stage least squares was used by Borders and Bailey(1986). In the compatible system of equations developed by Borders and Bailey(1986), some of the dependent variables from one equation appear as independent variables in other equations. The application of ordinary least squares in this system will result in biased and inconsistent parameter estimates due to the correlation between the independent variables and the error terms. The authors applied three-stage least squares to address this problem. Ordinary least squares was compared to unrestricted and restricted three-stage least squares. (Restricted meaning parameters that appear in more than one equation were constrained to be equal.) The parameter estimates were similar for all three cases, with three-stage least squares having generally smaller standard errors of prediction. Using restricted three-stage least squares so that the system will be compatible resulted in models that explained only slightly less variation than noncompatible ordinary least squares models.

Murphy(1983), started with a yield equation proposed by Schumacher(1939) and a modified version of the Chapman Richards growth function, developed a system of equations for computing projected basal areas, and current and future volumes for two merchantability classes. Murphy used seemingly unrelated nonlinear regression to fit the system. Seemingly unrelated nonlinear regression is an extension of the linear case developed by Zellner(1962). Since there is a strong likelihood of contemporaneously correlated residuals in Murphy's(1983) system, seemingly unrelated nonlinear regression would be a statistically sound method of estimating the coefficients. Comparison of using seemingly unrelated regression to fitting the system equation-by-equation with ordinary least squares did not show any improvement in the seemingly unrelated procedure.

There have been many papers published on growth and yield and the ones here are representative of the kind of work that has been done for whole stand growth and yield models with the concept of compatibility and/or numerical consistency considered. Other models compatible and noncompatible have been developed based on diameter distributions and individual trees, and a general review of growth and yield models can be found in in Burkhart(1981).

Chapter 3

Methods

DATA

The data used in this study were made available by the Virginia Division of Forestry. The data contain information from the remeasurement of permanent plots of loblolly pine established on old fields in the Virginia Piedmont and Coastal Plain.

Diameter at breast height for all trees was measured to the nearest inch, and the height of some trees was measured to the nearest foot. A height/diameter regression was estimated for each plot. Site index, base age 25 years, was predicted using a site index equation developed by Devan and Burkhart(1982) and cubic foot volume was predicted using individual tree volume equations published by Burkhart and others(1972).

Since periodic annual growth is being used as the instantaneous growth rate, and the relationship between basal area, age, volume, and growth is non-linear, a uniform remeasurement period will be used. The remeasurement period of five years occurred most frequently in the data, and therefore was selected as the period to be used. 105 observations were useable from the original data set. The data used, ranged in age from 12 to 40 years, in basal area from 50 to 166 square feet per acre, and in volume to a four inch top, assuming a .5 foot stump, from 800 to 4,615 cubic feet per acre. The site index, base age 25 years, ranged from 45 to 71.

Of the 105 observations, 70 observations were randomly selected to be used for fitting the equations. The remaining 35 observation were set aside to be used as a validation data set.

PROCEDURE

Three methods of estimating coefficients will be evaluated in this study. The system of equations used were based on the work of Clutter(1963).

METHOD 1

The first method is the estimation methodology used by Clutter. Clutter first postulated a volume yield equation [1] based on expected properties. A volume growth equation was develop by differentiating the yield equation with respect to age. The yield equation used is a whole stand equation for loblolly pine which predicts cubic feet per acre based on A = age, S = site index, and B = basal area/acre.

[1] $\ln(V) = \beta_0 + \beta_1 S + \beta_2 \ln(B) + \beta_3 A^{-1}$

β_0, β_1, β_2, and β_3 are regression coefficients.

The coefficients of the yield equation were estimated using linear regression. The volume growth rate obtained by differentiating equation [1] with respect to age, is

[2] $$\frac{\partial V}{\partial A} = \beta_2 V B^{-1} \frac{\partial B}{\partial A} - \beta_3 V A^{-2}$$

Since basal area is a function of age, when the yield equation was differentiated, the derivative of basal area with respect to age, $\frac{\partial B}{\partial A}$, is encountered. This term is the basal area growth rate. To find a basal area growth rate equation, the derivative with respect to age of a postulated basal area yield equation [3] was taken.

[3] $\ln(B) = c_0 + c_1 S + c_2 A^{-1} + c_3 \ln(D) A^{-1} + c_4 S A^{-1}$

D = basal area at age 20, c_0, c_1 ,... are regression coefficients.

The resulting equation [4] after differentiating and some algebraic manipulation is the basal area growth rate.

[4] $$\frac{\partial B}{\partial A} = -BLn(B)A^{-1} + c_0 BA^{-1} + c_1 BSA^{-1}$$

The coefficients of this equation were also estimated using linear regression. The basal area growth rate equation was then substituted into the the growth rate equation [2] to come up with the final growth rate equation [5].

$$[5] \quad \frac{\partial V}{\partial A} = -\beta_2 V Ln(B) A^{-1} + \beta_2 c_0 V A^{-1} + \beta_2 c_1 V S A^{-1} + \beta_3 V A^{-2}$$

At this point, instead of using the coefficients β and c estimated for the yield and basal area growth rate equations, the coefficients β_2, $\beta_2 c_0$, $\beta_2 c_1$, and β_3 of the growth rate equation were reestimated using linear regression. The result is an algebraically compatible growth and yield system, that is numerically incompatible due to coefficients which differ at different levels of the system.

Prediction equations for basal area and volume can be developed from the respective growth rate equations. Using the growth rate equation [5] a cubic-foot volume projection equation will be developed. First, a basal area projection equation is developed by integrating the basal area growth rate equation [4] from initial age and basal area to a projected age and basal area. Rewriting equation [4] and integrating

$$\int_{\beta_0}^{\beta_p} B^{-1}(c_0 + c_1 S - \ln(B))^{-1} \partial B = \int_{A_0}^{A_p} A^{-1} \partial A$$

then solving for β_p results in.

$$[6] \quad \ln(B_p) = c_0 + c_1 S - A_0 A_p^{-1}(c_0 + c_1 S - \ln(B_0))$$

A_p = projected age, A_0 = initial age, B_p = projected basal area, B_0 = initial basal area,

This equation can then be substituted into equation [5] for the ln(B). This substitution takes into account basal area being a function of age. Using the notation $B = B_p$ and $A = A_p$ bringing volume to the left side, and substituting.

$$\frac{1}{V}\frac{\partial V}{\partial A} = -\beta_2\left\{c_0 + c_1 S - A_0 A^{-1}(c_0 + c_1 S - \ln(B_0))\right\}A^{-1} + \beta_2 c_0 A^{-2} + \beta_2 c_1 S A^{-1} - \beta_3 A^{-2}$$

expanding and combining terms results in the following equation [7].

$$\frac{1}{V}\partial V = \{\left[(\beta_2 c_0 - \beta_2 \times c_0) + (\beta_2 c_1 - \beta_2 \times c_1)S\right]A^{-1} - [\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)]\left[A_0 A^{-2}\right] - \beta_3 A^{-2}\}\,\partial A \qquad [7]$$

The projected volume equation is derived by integrating equation [7] from initial age and volume to a predicted age and volume. The resulting equation can then be solved for predicted volume, giving a volume prediction equation based on initial age and volume.

$$\ln(V_p) = \ln(V_0) + [(\beta_2 c_0 - \beta_2 \times c_0) + (\beta_2 c_1 - \beta_2 \times c_1)\,S]\left[\ln(A_p) - \ln(A_0)\right] - A_0[\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)]\left[A_p^{-1} - A_0^{-1}\right] + \beta_3\left[A_p^{-1} - A_0^{-1}\right] \qquad [8]$$

The values for the β's and c's were found when equations 4 and 5 were fitted and can now be substituted in for the final volume prediction equation.

METHOD 2

This method is a modification of the previous method developed by Clutter. Developing the system first algebraically, before estimating any of the coefficients, will result in a slightly different system. The volume projection equation is different, and this form of the equation is the form needed for the system to be truly algebraically compatible.

Starting with the same volume yield equation

$$\ln(V) = \beta_0 + \beta_1 S + \beta_2 \ln(B) + \beta_3 A^{-1}$$

and the same basal area growth rate equation

$$\frac{\partial B}{\partial A} = -BLn(B)A^{-1} + c_0 BA^{-1} + c_1 SA^{-1}$$

as Clutter, the system will be developed algebraically. The volume growth rate equation is expressed as

$$\frac{\partial V}{\partial A} = \beta_2 V B^{-1} \frac{\partial B}{\partial A} - \beta_3 V A^{-2}$$

Substituting the basal area growth rate equation for $\frac{\partial B}{\partial A}$

$$\frac{\partial V}{\partial A} = -\beta_2 V Ln(B) A^{-1} + \beta_2 \times c_0 V A^{-1} + \beta_2 \times c_1 V S A^{-1} + \beta_3 V A^{-2}$$

Where these β's and c's are the same as those in the volume yield and basal area growth rate equations. The basal area prediction equation is then developed as in method 1 by integrating the growth rate equation.

$$\ln(B_p) = c_0 + c_1 S - A_0 A_p^{-1}(c_0 + c_1 S - \ln(B_0))$$

Substituting this in for the $\ln(B)$ in the volume growth rate equation, again letting B B_p and A A_p results in

$$\begin{aligned}\frac{1}{V}\frac{\partial V}{\partial A} = & -\beta_2\left\{c_0 + c_1 S - A_0 A^{-1}(c_0 + c_1 S - \ln(B_0))\right\}A^{-1} \\ & + \beta_2 \times c_0 A^{-2} + \beta_2 \times c_1 S\, A^{-1} - \beta_3 A^{-2}\end{aligned}$$

expanding and combining terms results in the following equation.

$$\frac{1}{V}\partial V = \left\{-[\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)][A_0 A^{-2}] - \beta_3 A^{-2}\right\}\partial A$$

The projected volume equation is then derived by integrating this equation from initial age and volume, to predicted age and volume. The resulting equation is then

$$\begin{aligned}\ln(V_p) = & \ln(V_0) - A_0[\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)][A_p^{-1} - A_0^{-1}] \\ & + \beta_3(A_p^{-1} - A_0^{-1})\end{aligned}$$

The system of equations is now

$$\ln(V) = \beta_0 + \beta_1 S + \beta_2 \ln(B) + \beta_3 A^{-1}$$

$$\frac{\partial B}{\partial A} = -BLn(B)A^{-1} + c_0BA^{-1} + c_1SA^{-1}$$

$$\frac{\partial V}{\partial A} = -\beta_2 V Ln(B) A^{-1} + \beta_2 \times c_0 V A^{-1} + \beta_2 \times c_1 V S A^{-1} + \beta_3 V A^{-2}$$

$$\ln(B_p) = c_0 + c_1 S - A_0 A_p^{-1}(c_0 + c_1 S - \ln(B_0))$$

$$\ln(V_p) = \ln(V_0) - A_0[\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)][A_p^{-1} - A_0^{-1}] + \beta_3[A_p^{-1} - A_0^{-1})$$

It can be seen that when the volume yield and the basal area growth rate equations are fitted, these estimates for the coefficients can be used throughout the system. Using these estimates will result in a system of equations that is numerically consistent as well as algebraically compatible.

METHOD 3

Although method 2 results in a numerically consistent set of equations, the estimates of the coefficients are not efficient estimators since they do not use all the available information. Using the same procedure as method 2 to develop the system, the equation of interest can be fitted using a simultaneous estimation procedure. The real use of the basal area and volume growth rate equations are to get to the prediction equations, and are of no practical use once the prediction equations are developed. The equations which will be used by a forest manager would be

$$\ln(V) = \beta_0 + \beta_1 S + \beta_2 \ln(B) + \beta_3 A^{-1}$$

$$\ln(B_p) = c_0 + c_1 S - A_0 A_p^{-1}(c_0 + c_1 S - \ln(B_0))$$

$$\ln(V_p) = \ln(V_0) - A_0[\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)][A_p^{-1} - A_0^{-1}] + \beta_3[A_p^{-1} - A_0^{-1})$$

The above equations can be fitted using seemingly unrelated regression described by Zellner(1962). Seemingly unrelated regression will allow all the available information to be used when the equations are fitted. The across equation coefficients constraints can also be applied, so the system will be numerically consistent as well as algebraically compatible. Using SAS's SYSNLIN (1984) procedure, the above three equations were fitted. If the growth rate equations are needed, the coefficients can be estimated by using the coefficients found in the three equations fitted, using the relationship between coefficients shown in method 2.

METHOD OF EVALUATION

The final step is to compare the predictive ability of the systems based upon the three methods of estimating coefficients: Method I, Clutter's procedure, where the yield equation, basal area growth rate equation, and the volume growth rate equations are fitted separately using ordinary least squares, and the coefficients from the basal area and volume growth rate equations are carried through the rest of the system; method 2, where the yield equation and the basal area growth rate equations are fitted using ordinary least squares, and these coefficients are carried through the rest of the system; method 3, where the yield, volume prediction and basal area prediction equations are fitted using seemingly unrelated regression. The equation coefficients for the three methods were estimated, and the effectiveness of each method was compared by looking at the ability to perform on independent data as well as how well they fit the data set the coefficients are estimated from. Several criteria were used when comparing the methods.

Calculated R^2 for the equations for the data which the equations where fit to was defined as

$$R^2 = \frac{\sum_{i=1}^{70} (y_i - \bar{y})^2 - \sum_{i=1}^{70} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{70} (y_i - \bar{y})^2}$$

This was calculated in terms of the measurement value of interest, for example, the yield equation is fitted in terms of logarithm of volume, but R^2 was calculated in terms of volume. Mean squared error (MSE) was calculated from the fitting data set.

$$MSE = \frac{\sum_{i=1}^{70} (y_i - \hat{y}_i)^2}{70}$$

For the 35 observations set aside as a validation data set the mean difference (MD)

$$MD = \frac{\sum_{i=1}^{35} (y_i - \hat{y})}{35}$$

and the mean absolute difference (MAD) were also calculate for each of the equations.

$$MAD = \frac{\sum_{i=1}^{35} /y_i - \hat{y}/}{35}$$

The percent difference (%)

$$\% = \frac{\sum_{i=1}^{35} (y_i - \hat{y})}{y_i}$$

and the percent difference on an absolute scale (/%/) were computed.

$$/ \% / = \frac{\sum_{i=1}^{35} /y_i - \hat{y}/}{y_i}$$

For each of the equations within each method, these criteria were compared.

Chapter 4

Results and Discussion

The coefficient estimates for the volume yield, volume prediction, and the basal area equations can be found in tables 1a, 1b, and 1c respectively. The estimates were obtained using the 70 observations, and the three methods described in the previous chapter.

For the yield equation

$$\ln(V) = \beta_0 + \beta_1 S + \beta_2 \ln(B) + \beta_3 A^{-1}$$

the estimates for each of the methods are very similar. Seemingly unrelated regression (SUR) results in estimates with smaller standard errors than the ordinary least squares (OLS) estimates of methods 1 and 2. Signs of the coefficients for this equation are what would be expected, positive relationships between volume and site, and volume and basal area, and a negative relationship with volume and inverse of age.

Table 1a. Coefficient estimates and their standard errors ( ) for the volume yield equation from the three methods.

Method	β_0	β_1	β_2	β_3
1	2.7243 (0.1009)	0.0153 (0.0011)	0.9874 (0.0223)	-14.1238 (0.4153)
2	2.7243 (0.1009)	0.0153 (0.0011)	0.9874 (0.0223)	-14.1238 (0.4153)
3	2.8035 (0.0883)	0.0145 (0.0010)	0.9809 (0.0197)	-14.1697 (0.3981)

Table 1b. Coefficient estimates and their standard errors ( ) for the basal area prediction equation from the three methods.

Method	c_0	c_1
1	4.7106 (0.5558)	0.0168 (0.0092)
2	4.7106 (0.5558)	0.0168 (0.0092)
3	3.7876 (0.5133)	0.0320 (0.0086)

Table 1c. Coefficient estimates and their standard errors ( ) for the volume prediction equation from the three methods.

Method	β_2	β_3	c_0	c_1	$\beta_2 c_0$	$\beta_2 c_1$
1	-0.3372 (0.4083)	0.1107 (7.4166)	4.7106 (0.5558)	0.01686 (0.0092)	3.2767 (1.7141)	-.00059 (0.0177)
2	0.9874 (0.0223)	-14.124 (0.4153)	4.7106 (0.5558)	0.01686 (0.0092)		
3	0.9809 (0.0197)	-14.170 (0.3981)	3.7876 (0.5133)	0.03199 (0.0086)		

For the basal area prediction equation

$$\ln(B_p) = c_0 + c_1 S - A_0 A_p^{-1}(c_0 + c_1 S - \ln(B_0))$$

the difference in coefficient estimates for method 3 from the other two methods are more pronounced then for the previous equation. The estimates of c_0 and c_1 for methods 1 and 2 come from the fitting of the basal area growth rate equation, while method 3 fits the actual basal area prediction equation.

The volume prediction equation is of the form

$$\begin{aligned}\ln(V_p) = {} & \ln(V_0) + \left[[(\beta_2 c_0 - \beta_2 \times c_0) + (\beta_2 c_1 - \beta_2 \times c_1)\, S]\left[\ln(A_p) - \ln(A_0)\right]\right]^1 \\ & - A_0[\beta_2 \times c_0 + \beta_2 \times c_0 S - \beta_2 \ln(B_0)]\left[A_p^{-1} - A_0^{-1}\right] \\ & + \beta_3\left[A_p^{-1} - A_0^{-1}\right]\end{aligned}$$

[1] This term is only present in method 1

Methods 2 and 3 produced similar coefficient estimates. Method 3 fitted the volume prediction equation directly, with the across equation restrictions applied, and method 2 used the coefficients from the volume yield and basal area growth rate equations. The first method uses the coefficients estimated in the volume growth rate and the basal area growth rate equations. The coefficients β_2, β_3, and $\beta_2 c_1$ have the opposite signs than one would expect in theory. In developing the system, the estimates for β_2 and β_3 in the volume prediction equation should equal β_2 and β_3 in the volume yield equation. The signs of these estimates and their magnitudes being wrong is the result of the fitting of the volume growth rate equation. There is high multicollinearity between the variables in the growth rate equation resulting in estimates that are wrong in sign, off in magnitude, and large in standard errors.

OLS and SUR give similar estimates for the yield and basal area prediction equations. Because of the multicollinearity, some of OLS estimates used in method 1, are of the wrong sign and have large standard errors, resulting in small t-values. The efficiency of the seemingly unrelated regression estimators relative to the ordinary least squares estimators can be determined by observing their standard errors, a smaller standard error would indicated a more efficient estimate. All coefficients show a slight increase in efficiency for method 3, which uses SUR to estimate the coefficients.

The three methods resulted in yield equations that fit the original data equally well, all R^2's and MSE's are very close (Table 2a). Based on the mean difference (MD), on independent data all three methods resulted in equations that over predicted volume, with method 3 having slightly more bias than the other methods. The mean absolute difference (MAD) shows method 3 to have slightly more variation than methods 1 and 2. On the percent scale the same results are true. All three methods produce equations that perform quite well.

The three basal area equations perform equally well in fitting the original data set (Table 2b). R^2 and MSE are almost the same for the three methods. All three equations slightly over predict, with method three showing slightly less bias. Method 3 again has the largest variation, but it is only slightly larger than the other two methods. Looking at the percent differences the same results are shown.

Table 2a. Evaluation of the volume yield equations.

	Method		
Procedure [1]	1	2	3
MD	-6.8518	-6.8518	-10.6330
MAD	48.9833	48.9833	54.0860
%	-0.6197	-0.6197	-0.8782
|%|	3.1900	3.1900	3.5344
R^2	0.9881	0.9881	0.9880
MSE	56.755	56.755	57.162

Table 2b. Evaluation of the basal area prediction equations.

	Method		
Procedure [1]	1	2	3
MD	-2.7210	-2.7210	-1.6662
MAD	5.9143	5.9143	6.3974
%	-3.3358	-3.3358	-2.2370
|%|	6.1737	6.1737	6.5334
R^2	0.8412	0.8412	0.8352
MSE	8.588	8.588	8.749

Table 2c. Evaluation of the volume prediction equations.

	Method		
Procedure [1]	1	2	3
MD	-74.1312	-55.2626	-33.3236
MAD	165.6653	181.4714	196.7352
%	-3.4804	-3.2722	-2.2417
|%|	7.9326	8.7299	9.3776
R^2	0.8138	0.8029	0.7920
MSE	270.296	278.057	285.673

[1] MD = mean difference, MAD = mean absolute difference, % = mean percent difference, /%/ = mean absolute percent difference, and MSE = mean squared error.

The volume prediction equations developed using the three methods also performed equally well in fitting the original data set (Table 2c). R^2 and MSE are almost the same for the three methods. All three equations over predict, with method 3 showing less bias. Method 3 again has the largest variation, but it is only slightly larger than the other two methods. Looking at the differences on a relative scale, the differences between the equations are minimal.

Examining the MSE for each method, method 1 is always less than or equal to method 2, and method 3 is always greater than method 2. Since there are no restrictions on method 1, it is logical that it would be able to minimize the sum of squared errors (SSE) better than method 2 and 3. The peculiar result is that method 2 is able to minimize the SSE better than method 3.

In method 2, the yield equation is fitted using OLS with no restriction applied, while for method 3 across equation restriction are applied. For this equation, the SSE for the OLS fit should be smaller, and table 2a shows this to be the case. Looking at the basal area prediction equation, coefficients for method 2 come from the fitting of the basal area growth rate equation, and the coefficients for method 3 come from the actual fitting of the basal area prediction equation with restriction on the coefficients applied. It is expected that fitting the actual prediction equation would result in a smaller SSE than using the coefficients estimated in basal area growth rate equation. This is not the case, since the across equation restrictions resulted in a poorer fit than using the growth rate coefficients, as can be seen by the MSE and R^2 in table 2b. The actual difference in the validation criteria is quite small.

Up until now, the procedures of method 1 and 2 are identical. Both these methods used unrestricted OLS to estimate the coefficients, while method 3 used SUR with across equation coefficient restrictions applied. The coefficient restrictions result in equations with slightly higher MSE, but it was hoped that this increase in MSE will be offset by a decrease in the MSE in the volume prediction equation.

In estimating the coefficients for the volume prediction equation, method 1 uses coefficients estimated when the volume growth rate equation was refitted. Using these estimates result in a numerically inconsistent system, but since there are no across equation restrictions imposed, this volume prediction equation has the smallest SSE. Method 2 fitted the volume yield and basal area growth rate equations, and used these coefficients in the volume prediction equation, resulting in a numerically consistent system. Since the coefficients are restricted to have the relationships between them needed for numerical consistency, the volume prediction equation's SSE is expected to be larger than the unrestricted procedure of method 1, and from table 2c it can be seen this is the case. Method 3 fits the volume prediction equation simultaneously with the volume yield and basal area prediction equations using SUR, with restrictions across equations which keep the system numerically consistent. It was thought that this procedure would be able to reduce the SSE in the volume prediction equation to counter the increase of SSE in the volume yield and basal area prediction equations. Instead of the decrease a slight increase in SSE is observed.

To find out if there are any particular areas in which any of the equations behave poorly, increasing the SSE, plots of the residuals $(y_i - \hat{y}_i)$, where y_i is the observed value and $\hat{y}_i$. is the predicted value, were examined. It was thought that placing restrictions on parameters may cause an equation to perform poorly in some regions of the data contributing significantly to the SSE.

Figures 1 and 2 are plots of the residuals from the volume yield equation. The residuals are plotted against the actual volume to see if there are any trends in the prediction, i.e. as volume increases residuals increase, etc. Both methods 1 and 2 are represented by figure 1 since they both use the same coefficient estimates. Figure 2 is the residual plot for the equation from method 3. No trends in the residuals are shown by these graphs.

The basal area prediction equation is also the same for methods 1 and 2, and their residuals are plotted on figure 3. The residuals from the equation for method 3 are plotted on figure 4. The residuals are plotted against the average basal per acre to look for trends with changes in basal area. The plots do not show any trends in the prediction of future basal area with changes in basal area.

Figures 5, 6, and 7 are the plots of the residuals for the volume projection equation from methods 1, 2, and 3, respectively. The residuals are plotted against average volume per acre. The residuals are randomly scattered about zero, showing no trends.

The residual plots for all the equations, from each of the methods, show no trends. The residuals fluctuate randomly above and below zero throughout the range of the data. There were no areas in the data that had any concentration of residuals to either side of zero, and no patterns are observed. This indicates that there is no problems in the model specifications and the restrictions on the parameters are not causing the equations to perform poorly within the range of the data.

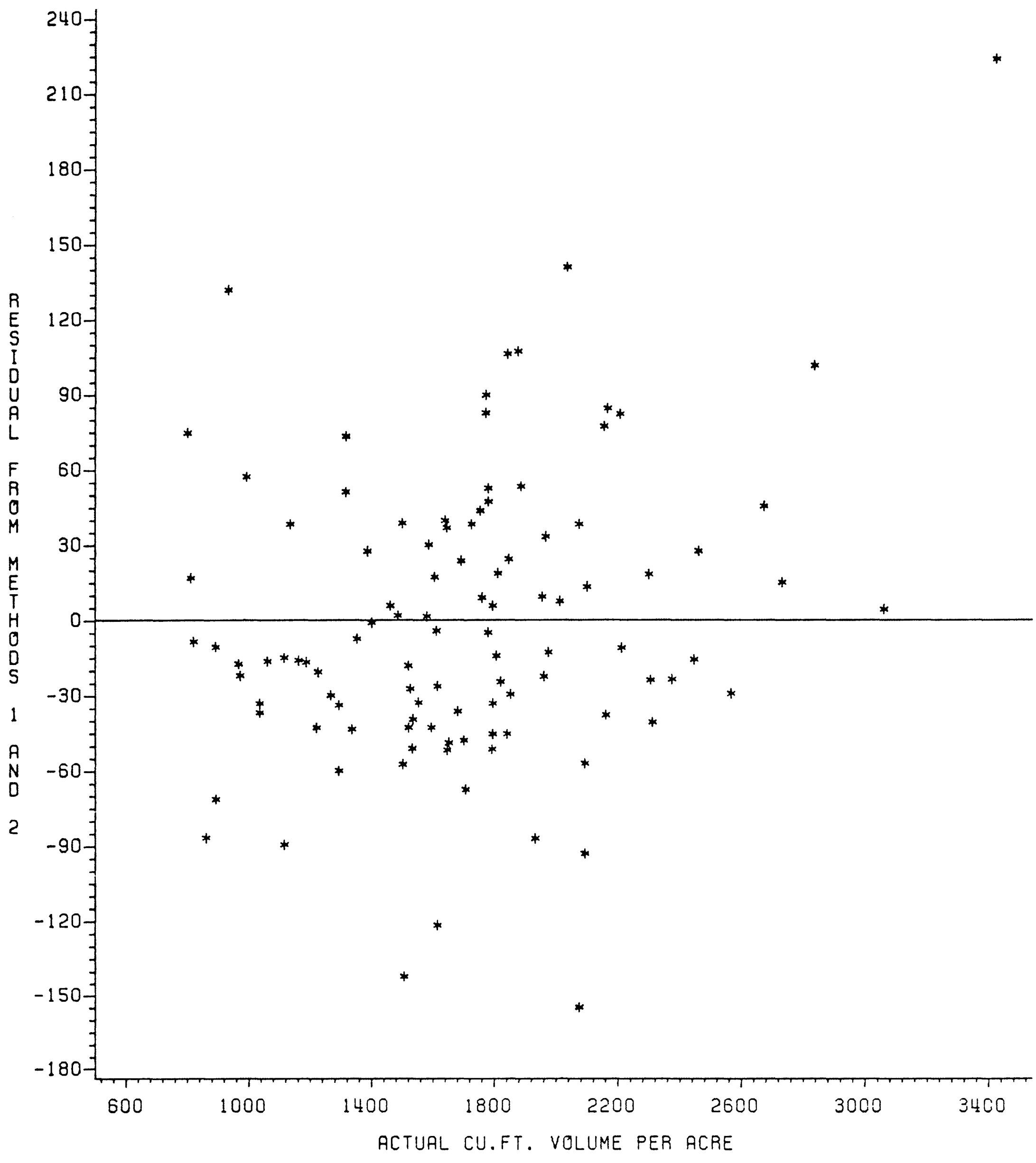

Figure 1. Plot of the residuals versus cubic foot volume per acre for the yield equation for methods 1 and 2.

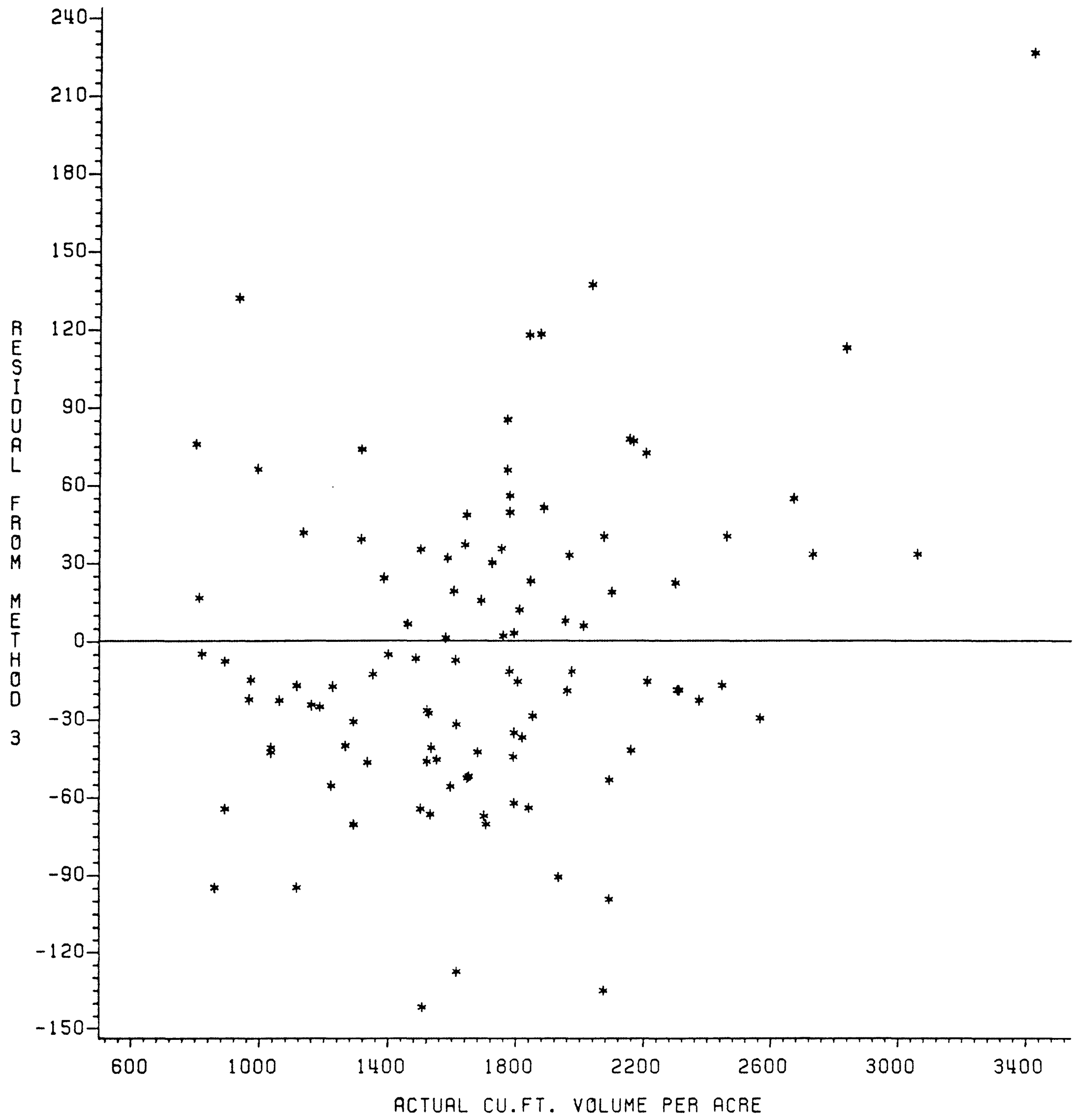

Figure 2. Plot of the residuals versus cubic foot volume per acre for the yield equation for method 3.

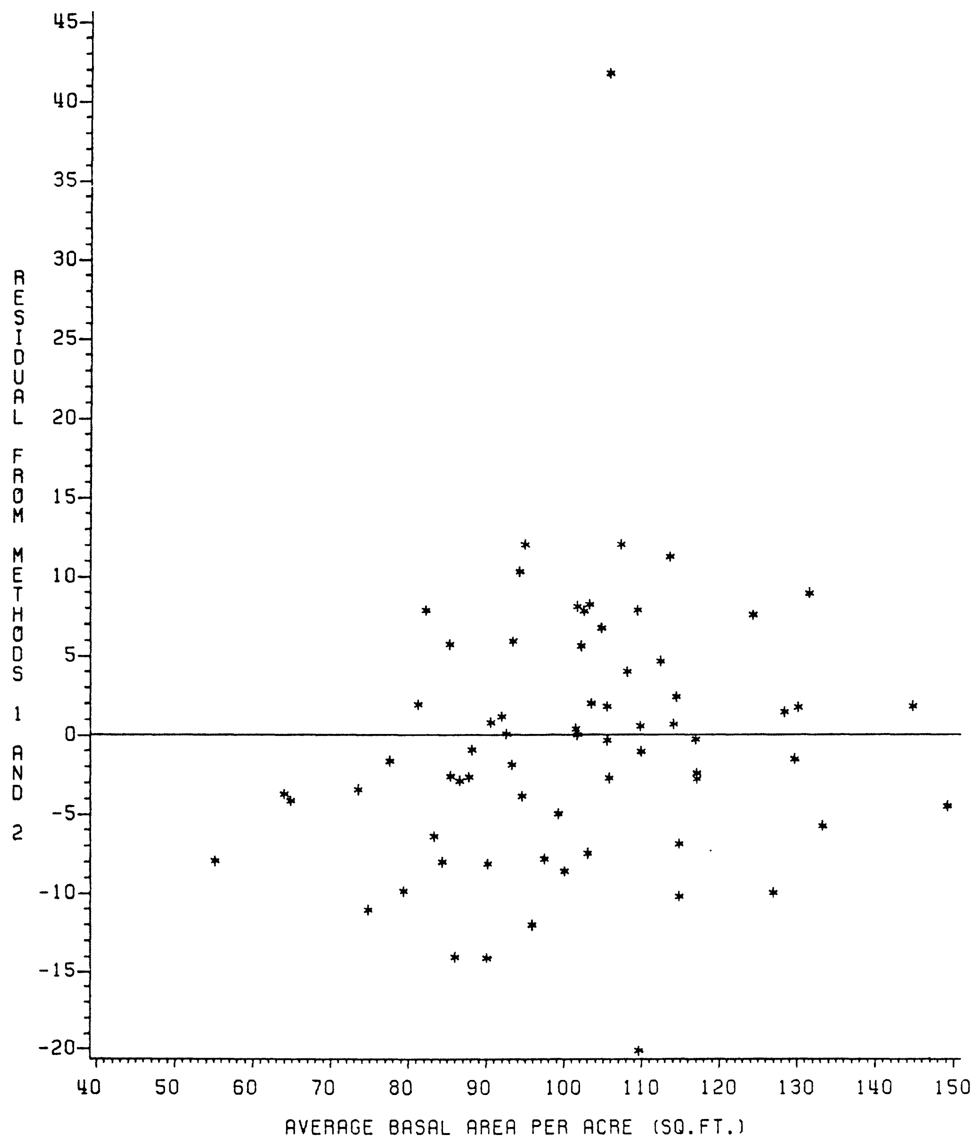

Figure 3. Plot of the residuals versus average basal area per acre for the basal area prediction equation for methods 1 and 2.

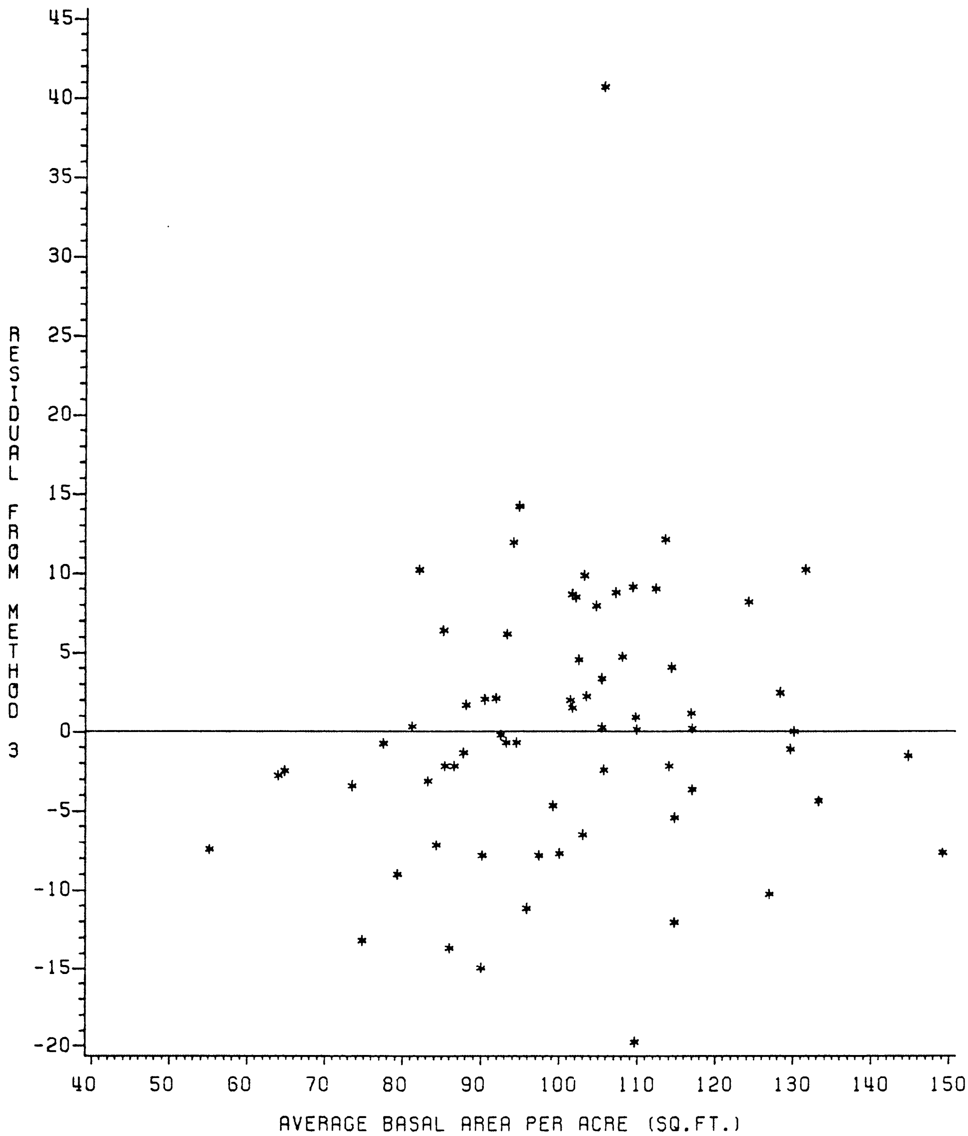

Figure 4. Plot of the residuals versus average basal area per acre for the basal area prediction equation for method 3.

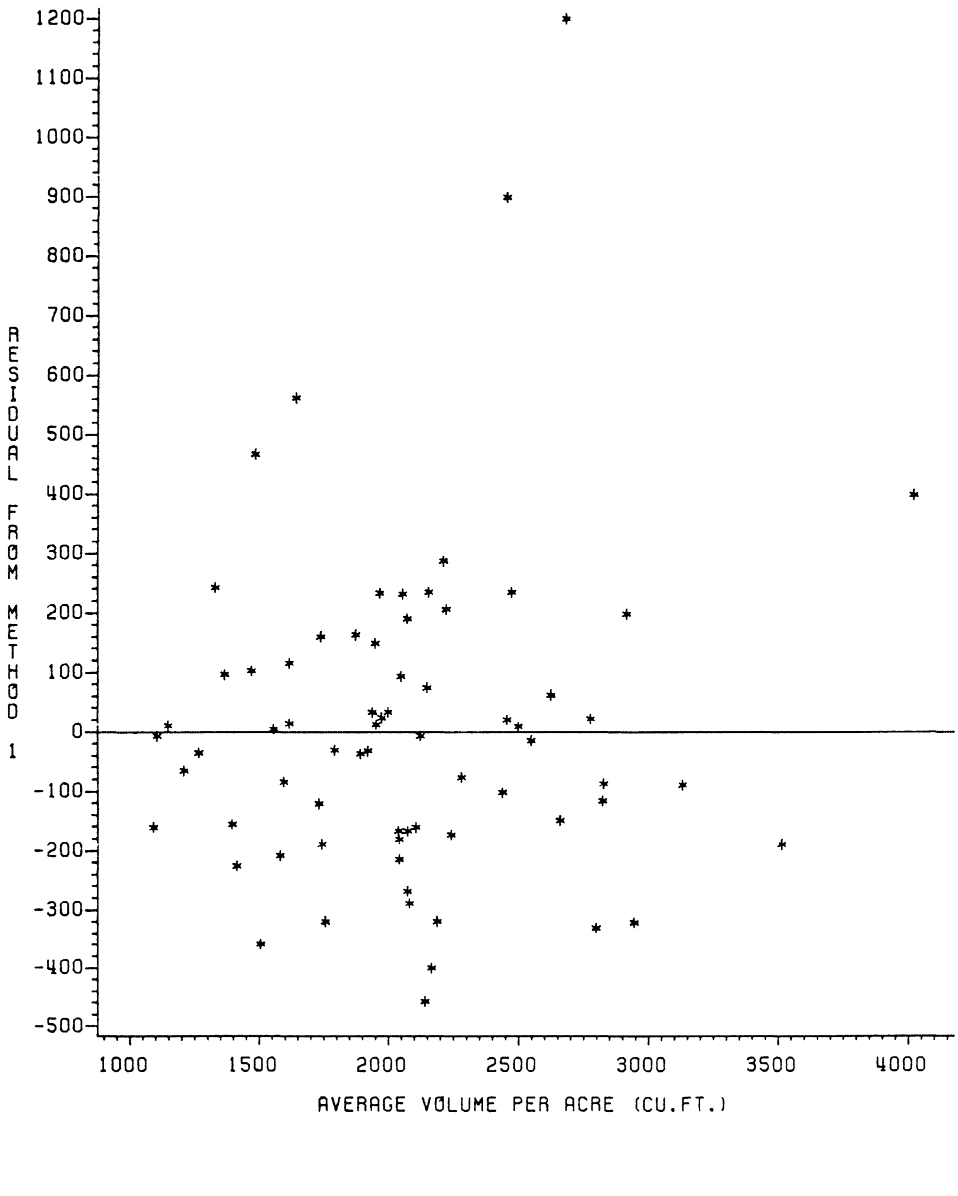

Figure 5. Plot of the residuals versus average cubic foot volume per acre for the volume prediction equation for method 1.

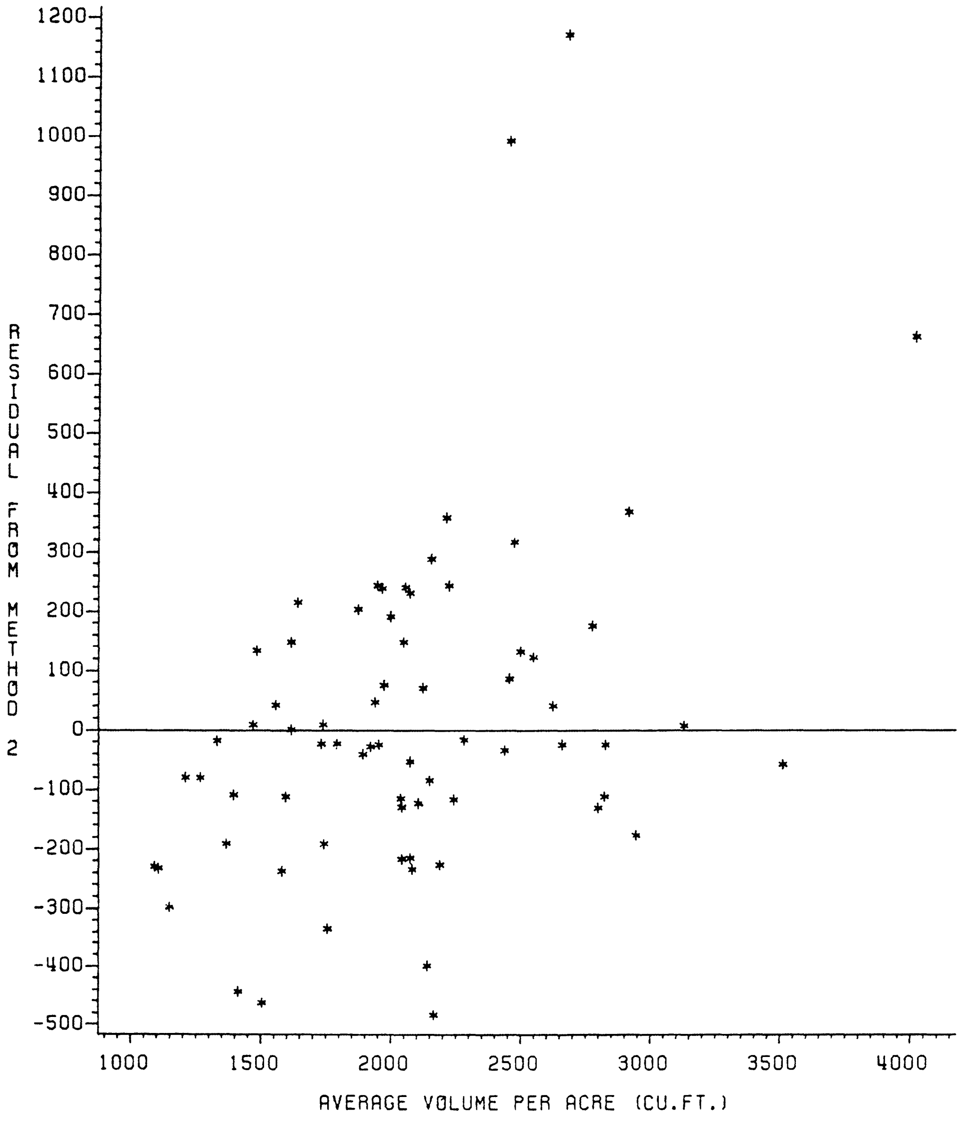

Figure 6. Plot of the residuals versus average cubic foot volume per acre for the volume prediction equation for method 2.

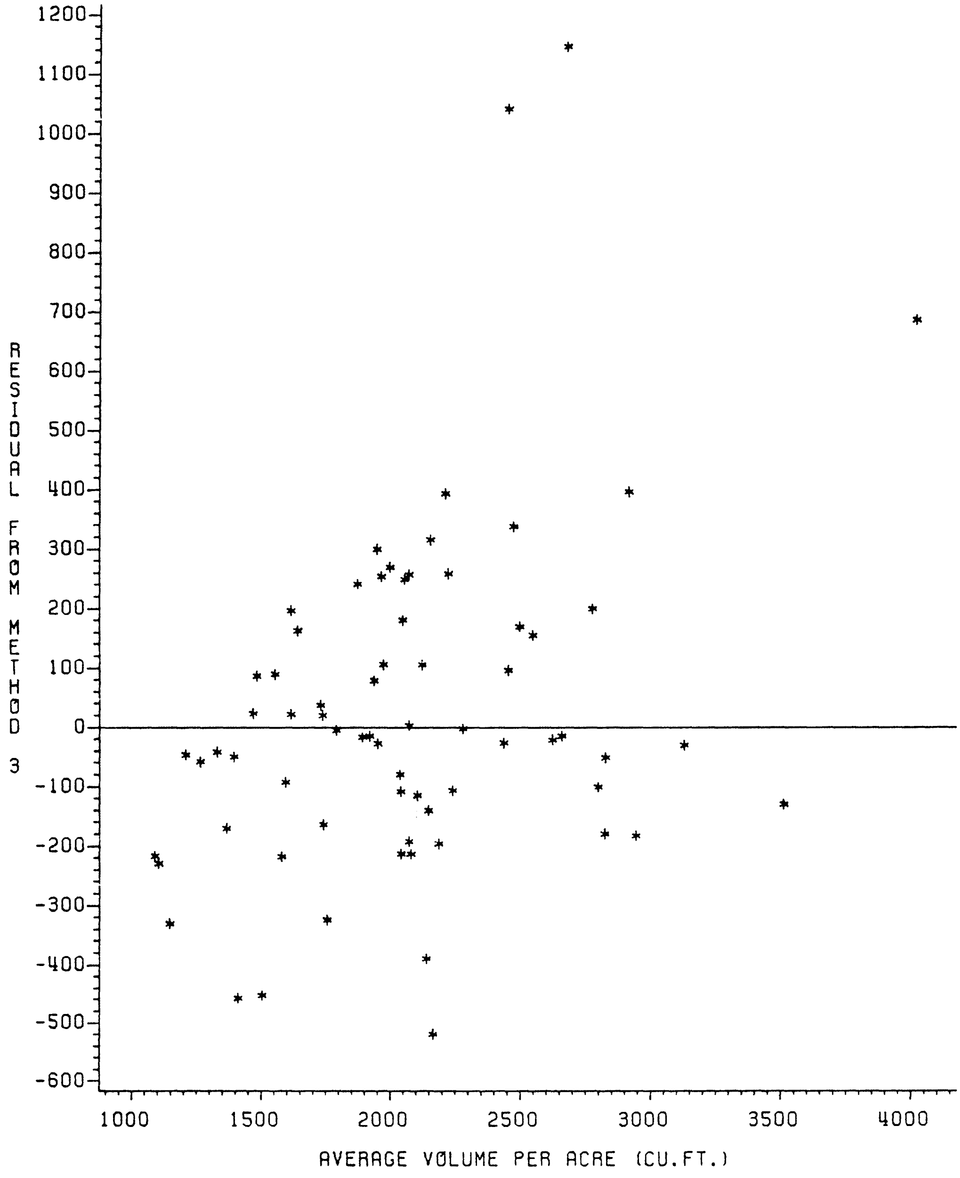

Figure 7. Plot of the residuals versus average cubic foot volume per acre for the volume prediction equation for method 3.

Looking at the SSE for each equation individually may not be appropriate in comparing the performance of the systems. Since interest lies in the system as a whole, not in the individual equations, some method of comparing the systems as one unit is needed. The straight sum of the SSE's from each of the equations for each method does not give equal weight to the equations since they are in different units. Dividing each SSE by an estimate of the variance for each equation is a way to eliminate the scale difference between the equations, but this measure of total system error does not incorporate any information about the correlation between the equations in the system. The reason for using SUR is because there exists correlation between the error terms in the system of equations and this information should be used when evaluating the system. Therefore the total system error (TSE) will be defined as

$$TSE = \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} \begin{bmatrix} \sigma_{1,1}^{-2} & \sigma_{1,2}^{-2} & \sigma_{1,3}^{-2} \\ \sigma_{2,1}^{-2} & \sigma_{2,2}^{-2} & \sigma_{2,3}^{-2} \\ \sigma_{3,1}^{-2} & \sigma_{3,2}^{-2} & \sigma_{3,3}^{-2} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where

e_i is a vector of errors, defined as $(y_i - \hat{y}_i)$, from the i^{th} equation.

$\sigma_{i,j}$ is the covariance between the i^{th} and j^{th} equations

$i = 1,2,3 \ \ j = 1,2,3$

equation 1 = volume yield

equation 2 = basal area prediction

equation 3 = volume prediction

This value will be a scalar that takes into account the fact that the errors are not independent by including cross products of the errors between equations divided by their covariances along with the SSE for each equation divided by its variance. The values of the variances and covariances used were estimated by fitting the system in its nonlinear form since the errors are defined in terms of volume and basal area, not logarithms of these values. Looking at the residuals plotted against actual values for the nonlinear equations showed no evidence of a violation of the homogeneity of variance assumption needed to use these estimates. Using the variances and covariances estimated by SAS's SYSNLIN procedure for the 70 observations to which the equations were fitted, the TSE was calculated for the each of the methods on the validation data set.

Table 3 shows the values for the total system errors for each method. When the information on the correlation between the equations within the system is taken into account, the usefulness of the SUR procedure can be seen. The total system error for method 3 is smaller than either of the other two methods.

Table 3. Total system error for the three methods.

Method	TSE
1	213.3198
2	237.9079
3	182.6572

Chapter 5

Conclusion

Clutter's (1963) system of growth and yield equations was analyzed, and it was found that that by fitting the volume growth rate equation without constraining the coefficients to have the relationships they need with the volume yield and basal area growth rate equations to be numerically consistent, the algebraic compatibility of the system is also lost. The system was developed algebraically without fitting any of the equations. The resulting system has a different volume projection equation than the one proposed by Clutter(1963). If the system is then fitted with across equation constraints on the coefficients, the system will be numerically consistent and algebraically compatible.

Two methods of estimating the coefficients in the numerically consistent system, and Clutter's original method were examined. Clutter's method, and one of the numerically consistent methods used ordinary least squares(OLS) to fit part of the systems equations independently, and then use these estimates for the coefficients in the rest of the system. The third method fitted the system simultaneously using seemingly unrelated regressions (SUR) with constraints on the coefficients to keep the system consistent.

Looking at the system equation by equation, Clutter's unconstrained equations fit the data best, based on R^2 and MSE values. On independent data the performance of the methods varied. Examining the mean difference and the mean absolute difference, the SUR procedure tended to have slightly less bias, but slightly more variation. On the relative scale very little difference between the methods is apparent. Examining the t-values, the SUR procedure usually produced the most efficient estimators, although the gain in efficiency does not appear large.

Looking at the system as a whole, which is important when examining a compatible system of growth and yield equations, the SUR procedure should be given consideration when estimating the coefficients. Although independent OLS fits to each equation cannot be beaten for the individual equation, there are other aspects that must be considered. An assumption made when using OLS is that there exists no other model that is correlated to the error term in the postulated model. This assumption is violated in many growth and yield systems. One has to consider that volume, volume growth, and basal area growth are not independent of each other. SUR is one method of estimating the coefficients of equations that have correlation between their error terms. In "classical" SUR as defined by Zellner(1962) the correlation of the error terms is due to missing independent variables, and there is no functional relationship between the dependent variables contributing to the correlation. In the system of equation used in this paper, there exists functional relationships between the dependent variables. It is unknown what the effects this type of violation has on the estimation procedure, but since SUR was able to improve the efficiency of the coefficient estimates in this paper, the method appears to be a viable alternative to OLS. When a measure of total system error which accounts for the equations not being independent was examined, the SUR procedure is found to be superior to the constrained and unconstrained OLS methods.

SUR provides a convenient method of fitting a system of growth and yield equations that may have correlation in residuals, and allows for the across equation restrictions that may be needed for consistent estimates. In this paper using SUR resulted in more efficient coefficient estimates, and a reduction in total system error. Although the system does not fall exactly in the case described by Zellner(1962), using SUR resulted in improved estimates.

Literature Cited

Borders, R.E. and R.L. Bailey 1986. A Compatible System of Growth and Yield Equations for Slash Pine Fitted with Restricted Three-Stage Least Squares. For. Sci. 32(1):185-201

Buckman, R.E. 1962. Growth and Yield of Red Pine in Minnesota. U.S. Dept Agr Tech Bull 1272, 50pp.

Burkhart, H.E. 1981. State of the Art in Predicting Growth and Yield. Proceedings Eleventh Forestry and Wildlife Forum (R.L. McElwee,Ed.). School of Forestry and Wildlife Resources, VPI & SU Publication 420-001:19-36.

Burkhart, H.E., R.C. Parker, M.R. Strub, and R.G. Oderwald. 1972. Yieldsof old field loblolly pine plantations. Div For and Wildl Resour, Va Polytech Inst and State Univ. Publ FWS-3-72, 51p.

Burkhart, H.E. and P.T. Sprinz. 1984. Compatible Cubic Volume and Basal Area Projection Equations for Thinned Old-field Loblolly Pine Plantations. For.Sci. 30(1):86-93

Clutter, J.L. 1963. Compatible Growth and Yield Equations for Loblolly Pine. For. Sci. 9(3):354-371

Curtis, R.O. 1967. A Method of Estimation of Gross Yield of Douglas-fir. Forest Science Monograph 13 24pp.

Devan, J.S., and H.E. Burkhart. 1982. Polymorphic Site Index Equations for Loblolly Pine Based on a Segmented Polynomial Differential Model. For. Sci. 28:544-555

Moser, J.W. Jr., and O.F. Hall. 1969. Deriving Growth and Yield Functions for Uneven-aged Forest Stands. For.Sci. 15(2):183-188

Murphy, P.A. 1983. A Nonlinear Timber Yield Equation System for Loblolly Pine. For. Sci. 29(3):582-591

SAS Institute. 1984. SAS/ETS User's Guide, Version 5 Edition. SAS Institute, Cary, NC, 505-550

Schumacher F.X. 1939. A New Growth Curve and its Application to Timber-Yield Studies. J. Forestry. 37:819-820

Smith V.G. 1983. Compatible Basal Area Growth and Yield Models Consistent with Forest Growth Theory. For. Sci. 29(2):279-288

Sullivan A.D. and J.L. Clutter. 1972. A Simultaneous Growth and Yield Model for Loblolly Pine. For. Sci. 18(1):76-86

Zellner, A. 1962. An Effective Method for Estimating Unrelated Regressions and Tests for Aggregation Bias. J. Am. Statist. Assoc. 57:348-368

The vita has been removed from the scanned document