AN ABSTRACT OF THE THESIS OF

Jeff D. Hamann for the degree of Master of Science in Forest Resources presented on December 13, 2002.

Title: Simultaneous Equation Estimation for Individual Tree Growth in Young Southern Oregon and Northern California Conifer Plantations

Signature redacted for privacy.

Abstract approved:

David D. Marshall

This thesis presents methods for obtaining asymptotically efficient and consistent parameters and variance estimates for simultaneous equations in a forest growth modelling context. Ordinary Least Squares (OLS), Seemingly Unrelated Regressions (SUR), Two-Stage Least Squares (2SLS) and Three-Stage Least Squares (3SLS) are presented for linear models. The variables, model types and transformations are examined for appropriateness in diameter and height growth models in young stands. A basal diameter growth, height growth and static crown ratio model were developed using the methods described. Model performance was measured by the ratio of the standard-errors of the predictions for basal diameter growth, height growth and crown ratio as described by Hasenauer et al. (1998). The 3SLS model performed better than the 2SLS or OLS for the basal diameter growth. The advantages of using 3SLS over 2SLS or OLS for the height growth and crown ratio models were minimal. Finally, a simultaneous equation estimation package was developed for the R (Ihaka and Gentleman, 1996) open-source computer program.

Simultaneous Equation Estimation for Individual Tree Growth in Young Southern Oregon and Northern California Conifer Plantations

by

Jeff D. Hamann

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented December 13, 2002
Commencement June 2003

ACKNOWLEDGMENTS

Without the contribution of a number of people, this work would have not been possible. I would like to thank Wanda Sinicope for her unwavering support, wisdom and grandmotherly love. I would like to thank Martin Ritchie from the Pacific Southwest Research Station for having the vision and fortitude to fund the work, providing the data and for being a good friend. I would also like to thank David Marshall for allowing me to continue fully funded and unsupervised, Doug Maguire for giving me the opportunity to shanghi him for my committee, Darius Adams for being in the wrong time at the right place and David Brauner for being my Graduate Council Representative.

I would also like to take this opportunity to thank those friends that helped me along the way in life. I would like to extend my thanks to Don Wallace for having the patience to listen even when I had nothing to say and for the wisdom to let me learn from my mistakes. I owe a special thanks to Fred Nicoll for giving me a first chance; and then a second. I would like to thank David Walters and Greg Johnson for letting me develop my skills to support myself. As well, I would like to thank John Sessions for presenting the opportunities to make use of those skills. And I would like to thank Adam Lewenberg, who was a great help in preparing the formatting.

I would finally like to thank the people in my life that have made this journey at times a wonderful experience and at others, a living hell. You know who you are.

DEDICATION

For Wanda and Mary.

I couldn't have made it without you.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure | Page

LIST OF TABLES

Table | Page

1. INTRODUCTION TO INDIVIDUAL TREE ESTABLISHMENT GROWTH AND YIELD MODELS

1.1 ABSTRACT

The paper presents a brief summary of forest growth and yield models for young stands in Southern Oregon and Northern California. The focus is on single-tree density-dependent models. The variables, transformations, model forms, fitting methods and regression diagnostics are presented as a background for the development of a system of simultaneous equations to estimate diameter increment and height increment in Southern Oregon and Northern California for Douglas-fir (*Psuedotsuga menzesii* (Mirb.) Franco). A static crown ratio model was included in the system of equations for use in model diagnostics and for projection purposes.

1.2 INTRODUCTION

Growth models help resource planners determine future forest conditions, estimate potential harvest levels, and simulate new silvicultural techniques without costly and long experimentation. Models are available for various aspects of forest growth at different resolutions. Whole stand models project stand level attributes and are commonly used in large scale planning projects where individual tree attributes are unavailable for each modeling unit; typically a stand polygon (Curtis et al., 1982). Yield curves are available to estimate volumes or simple stand attributes when simulating silvicultural treatments is not required. Some yield curves contain the ability to address variable stand densities so that thinning stands over time can be simulated (Chambers, 1980). Succession or process models are often used to study population or forest gap dynamics by modeling atmospheric, nutrient, and photosynthate allocation (Botkin et al., 1972). Inventory projection models, at the stand or tree level, are designed to update estimates of forest inventories over short periods of time and are often extended to simulate silvicultural treatments in older stands (Hann et al., 1993; Stage, 1973a; Wensel et al., 1987).

In the Pacific Northwest, inventory models are typically developed as single tree models. These models include a set of equations to estimate the increments of different tree attributes. These equations for individual trees estimate diameter growth at breast height (DBH), total height growth, crown recession and the probability of mortality (Hann et al., 1993; Wensel et al., 1986; Wykoff et al., 1982). These models are intended for use on established stands where an inventory sample provides a tree list that is projected forward under various silvicultural regimes and are not considered to be in the domain of stands in the initiation phase.

Growth models for young stands contain unique problems for forest growth and yield modelers. Stand dynamics are somewhat simplified because crown recession has not yet occurred, thus crown ratio dynamics are less influential. Many of the trees are still below breast height when first measured, thus modeling diameter increment at breast height is often irrelevant. In these models, the goal is to generate a tree list that is applicable to those stands found in the lower range of inventory update models. In some cases, young stand models have been developed so that they are robust enough to overlap with the lower age range of inventory update models (Ritchie and Powers, 1993).

Growth models are typically built as a set of equations applied to the same modeling unit; a single tree. Often these component models are constructed from different datasets drawn from the same population. For example, a diameter increment model may use diameter increments from the same trees as the height increment model, a subset or a superset. Regardless, any information pertaining to the data that was used in the height increment model is ignored. The fact that the errors in the predictions are likely to be correlated has been largely ignored until recently (Furnival and Wilson, 1971; Tang et al., 2001).

The relationship between the two equations may change over time as well. As diameter growth and height growth are seemingly independent for a single tree, the ratio of height growth and diameter growth may change with respect to species, site quality, competition, and time.

This chapter presents a literature review of single-tree diameter and height growth models that have been developed for young plantations in Southern Oregon and Northern California. There are two sections detailing the independent variables for diameter increment and height increment, a brief presentation of model types and a final presentation of linear and non-linear model forms.

1.3 EXPLANATORY VARIABLES

The explanatory variables used to develop increment models for trees have been developed to relate simple field measurements to stand structures and stand dynamics. They should convey information without obfuscation as well as describe the relationships among other dependent variables and the other components of a growth and yield system. Independent variables for diameter and height increment equations may include age, variables describing plant size, vigor, competiton.

1.3.1 Diameter Growth

Diameter growth is one of two primary variables predicted in estimating tree growth in young stands. The diameter growth of a single stem can be modeled as radial increment, diameter increment, basal area increment, future diameter, or future basal area (Vanclay, 1994). Since all are related mathematically it has been shown that there is no difference in fitting one over the other (West, 1980; Larsen and Hann, 1987); hence this section will focus on diameter increment.

Regressor variables in diameter growth equations typically include plant size, vigor, competitive position, population density, and site productivity (Wykoff, 1990; Vanclay, 1994; Martin and Ek, 1984). A summary of how various plant attributes have been used in diameter growth equations is presented in Table 1.1.

1.3.1.1 Plant Size

Simple field measurements are often used as measures of tree size. Stem diameters (basal diameter at 6 inches above the root collar (D6) for small trees and at breast height (DBH) for larger or older trees) and total height (THT) and their transformations are the most common when predicting diameter increment (Ek and Monserud, 1974; Dolph, 1988b, 1992b; Ritchie and Hann, 1986; Arney, 1985; Wensel et al., 1986; Hann and Ritchie, 1988; Opalach et al., 1990; Ritchie and Powers, 1993; Hann and Larsen, 1991; Wykoff et al., 1982; Wykoff, 1990).

Age, which has often been used to describe plant size in height growth models, can also be used to describe plant size in diameter increment models. While tree age in plantations may be a strong variable in model fit statistics, models developed from plantations with multiple plantings or severe browse damage may not predict individual trees well nor be as robust as a model that does not use individual tree age as a surrogate for plant size. Using age has additional disadvantages when compared to simple field measurements such as diameter and height when modeling older trees. Obtaining whorl counts on young trees is easy if crown closure has not occurred, otherwise, stem cores are required and both are time consuming and may be prone to error.

1.3.1.2 Vigor

Tree vigor attributes to describe a plant's ability to utilize resources, can be characterized by crown attributes, current growth rates, and previous growth rates. Crown ratio (CR) is often used to characterize tree vigor because of the inherent association with dominance and influence on photosynthetic capacity (Hann and Ritchie, 1988; Ritchie and Hann, 1985, 1986; Wensel et al., 1986; Wykoff, 1990). Hann and Larsen (1991) and Ritchie and Hann (1985) used initial crown ratio as a measure of vigor in the development of a potential-modifier function for predicting five-year diameter and basal area increment growth for species in Oregon, respectively.

Other authors have successfully used height growth ΔTHT in diameter increment models (Krumland and Wensel, 1981; Ritchie and Hann, 1985). While height growth may be a better indicator of young tree vigor, the authors failed to fit the height growth and diameter growth model simultaneously as a system of equations, thus ignoring the "errors-in-variables" problem. Ritchie and Hann (1985) rejected using Seemingly Unrelated Regressions (SUR) (Zellner, 1962) due to the difficulties of fitting nonlinear regressions and because the data were very unbalanced with respect to height growth and diameter growth measurements.

1.3.1.3 Competitive Position

The competitive position of a tree should relate information about how a tree will respond to a sudden change in the competitive position, such as a release due to thinning, as well as the tree's ability to change competitive position when grown in a stand (Vanclay, 1994). Simply, this measure represents the ability of the tree to advance past cohorts or remain in the same position when unaffected by competition.

In diameter increment models, a tree's ability to compete is related to it's ability to adjust to environmental changes within the stand or modify their environment to their advantage and to the detriment of other trees within the stand (Vanclay, 1994). An example would be in the method of competition between two common forms of trees, hardwood or broadleaf and conifers. Hardwood trees often compete "out" or try to limit the amount of sunlight reaching the needle bearing trees until the hardwood species can establish dominance whereas the conifers attempt to grow taller than the hardwood stems to assert dominance during early stages of stand development. To capture these attributes, some index or measure of the competitive position is typically included in diameter increment models.

Some researchers have suggested that competition in plantations is primarily for light resources which would suggest that competition from taller trees is "one sided". That is to say, shorter trees or more precisely, lower leaves are affected by leaves higher in the canopy (Vanclay, 1994). Some authors suggest that using basal area in larger stems (BAL) complements stand basal area (SBA) in diameter increment functions (Cannell et al., 1984; Ford and Diggle, 1981). A rationalization for this conclusion is that basal area is

Reference	Size	Vigor	Competitive Position (One-Sided)	Density (Two-Sided)	Site Productivity
Hann and Larsen (1991)	DBH	CR	BAL	SBA	Site Index
Ritchie and Hann (1985)	DBH	CR, ΔTHT	$CCFL_{plot}$	SBA	
Ritchie and Hann (1985)	DBH	CR	$CCFL_{plot}$	SBA	Site Index
Uzoh et al. (1998)	DBH	N/A	$\frac{BAL}{DBH}$		
Dolph (1988b)	DBH	CR	BAL	SBA	Elevation Slope Site Index
Dolph (1992b)	DBH	CR	BAL	SBA	Slope Aspect Site Index
Huang and Titus (1995)	DBH THT	N/A	$\frac{DBH_i}{QMD}$	SBA TPA	SPI
Nystrom and Kexi (1997)	DBH Age BAL	CI	RDBH	SBA	Site Index
Wykoff (1990)	DBH	CR	BAL	CCF	Slope Aspect Elevation
Smith and Bell (1983)	DBH	N/A	CSI, ΔCSI	N/A	N/A
Donnelly (1997)	THT	CR	N/A	N/A	N/A
Donnelly (1997)	DBH	CR	BAL RTHT	SBA CCF	Site Index Elevation Slope Aspect Location
Krumland and Wensel (1981)	DBH THT	CL ΔTHT	CC_{66}	N/A	Site Index

Table. 1.1: Common explanatory variables used in diameter increment models

highly correlated to sapwood area which in turn is highly correlated to leaf area (Waring et al., 1980). The leaf area above a tree may be the most direct measure of the tree's competitive position, but obtaining precise measurements for model development are problematic.

Modeling the interaction between tree size and BAL can give better results than using BAL alone (Wykoff, 1990). Wykoff (1990) also concluded that using BAL was more appropriate than using a measure of relative size because relative tree size ignored the influence of density management practices (Vanclay, 1994).

The competitive influence of other species has been examined by some authors when developing young tree models under the influence of shrub competition (Opalach et al., 1990; Szwagrzyk, 1997). This robust approach has advantages by estimating the species specific competitive influence on a subject tree. The disadvantage to using a competitive index based on species is found in the non-quantitative nature of competition and the difficulty of measuring the competitive influence of various species especially shrub or herbaceous competition (Szwagrzyk, 1997; Burton, 1993; Wagner and Radosevich, 1991b,a; Brand, 1986). Opalach et al. (1990) found fitting small datasets and qualitative measures of species specific competition indicies difficult.

Spatial information has been examined by some authors (Ford and Diggle, 1981; Burton, 1993; Brand, 1986). Brand (1986) examined the relationship between horizontal proximity of competing vegetation and the influence on planted Dougals-fir in southwestern British Columbia. The author found the best relationship was between shrub canopy height and the height of the subject plant. Burton (1993) suggested using neighborhood competition indices that included temporal relationships, citing the example that the growth rate of trees under demonstrable competition, is still greater than the surrounding shrubs and will eventually escape from the competitive constraints of the surrounding shrubs. Spatial indices may not be practical in forest management due to the economics of obtaining the required data (Vanclay, 1994).

Other competitive position measures have been expressed as well. Uzoh et al. (1998) used a ratio of BAL/DBH as a measure of competitive position with promising results. The authors found that this measure reflected those results produced by others (Wykoff, 1990), where the parameter estimate had a negative coefficient suggesting that diameter increment should decrease as the value of BAL/DBH increases, and the magnitude would suggest the competitive status relative to other trees on a plot. Huang and Titus (1995) included a ratio of the subject tree species total stand basal area to the total stand basal area. The authors found that the term was negative, suggesting that for two spruce-aspen stands; the stand with more spruce will have lower diameter growth rates. They speculated that the ratio may not be directly related to interspecific competition and stands dominated by aspen may contain healthy spruce trees with large diameters. Another relative measure used by Nystrom and Kexi (1997), was the inverse of the index used by Uzoh et al. (1998).

To include an index measure that was sensitive to treatments and time, Smith and Bell (1983) examined both initial competitive stress index (CSI) (Arney, 1973) and the change in CSI and found that including both CSI and ΔCSI gave superior results to diameter increment models than including only a single measure. Smith and Bell (1983) defined CSI as,

$$CSI_j = 100\frac{\Sigma AO_{ij} + A_j}{A_j} \tag{1.1}$$

where CSI_j is the competitive stress index of the jth subject tree, AO_{ij} is the area of overlap of the ith competitor's growing space with that of the subject tree and A_j is the growing-space area of the jth tree. A CSI value of 100 represents fully occupied growing-space by the subject tree and a value of 200 means an additional 100 percent of the growing-space is occupied by the growing-space area of its neighbors. The ΔCSI term was included to reflect changes in basal area due to treatments whereas the static CSI measure reflected the current competitive conditions.

The quantification of the competitive position for a tree is critical in the development of models. In short, the competitive position dictates how well a tree is able to compete against other trees and is usually conveyed in a single value or simple function such as CR or crown closure at the tip of the subject tree (CCH). In young stands, the most influential explanatory variables may be a function of height and diameter.

1.3.1.4 Density

Stand density, a measure of competition, influences diameter increment by limiting the amount of photosynthate generated by and allocated to each tree when stands approach the maximum size-density line (West et al., 1997; Enquist et al., 1998).

The relationship between density and diameter growth is useful because it is not related by age, site productivity, competition levels, or treatment histories and should provide an upper limit on the available space, allow for simple interpretation and follow allometric relationships (Curtis, 1982). Many measures of stand density have been developed from DBH such as relative density (Curtis, 1982) and stand density index (Reineke, 1933; Curtis, 1982). The maximum size-density line and the relationship to the self-thinning rate should be reflected in the diameter growth rates and vice-versa.

In diameter growth models, stand or plot basal area is typically used to describe density (Hann and Larsen, 1991; Dolph, 1988b, 1992b; Ritchie and Hann, 1985; Huang and Titus, 1995; Nystrom and Kexi, 1997). Stand basal area is used because the fundamental measurement, DBH, is easily obtainable and is directly related to the change in diameter. This makes it a simple and natural choice for a density measure.

Other measures of stand density have been used as well. Crown Competition Factor (CCF) was applied by Wykoff (1990) and Donnelly (1997) in the PROGNOSIS model for large trees. Crown Competition Factor (Krajicek et al., 1961) is defined as,

$$CCF = \sum_{i=1}^{n} (0.001803 * MCW_i^2 * EXPF_i) \tag{1.2}$$

where MCW is the maximum crown width, in feet, for the tree and $EXPF$ is the expansion factor, in stems·ac^{-1}, for the tree record. Crown competition factor was developed to express the sum of the potential crown area per unit area as a measure of "available open space". The authors speculated that the differences among maximum CCF values could be attributed to open grown tree crown development, basic crown shape, and species tolerance.

Huang and Titus (1995) included trees per hectare as well as SBA in their model to estimate annual diameter increment. Schroder and von Gadow (1999) used a relative spacing index (RSI) to describe the density of the stand in the development of a basal area increment function. The RSI was defined as,

$$RS_i = \frac{\sqrt{\frac{10000}{N}}}{H_D} \tag{1.3}$$

where N is the number of stems per hectare and H_D is the dominant stand height using an anamorphic height model. While the authors found the approach gave much better results than BAL, the inclusion of dominant height in a diameter growth model density measure is not appealing as it is prone to measurement error for age and height observations if the relationships among the other equations in a set of increment equations are not addressed.

1.3.1.5 Site Productivity

Site quality is perhaps the most influential category regarding increment functions because of the wide range of site productivity in the Pacific Northwest Region. To predict future conditions over a wide range of site qualities, it is necessary to obtain site quality estimates from either existing site quality attributes or from inherent site attributes. Both methods of describing site quality have inherent qualities and disadvantages and both have been used in diameter increment models.

The most common measure of site quality is site index. Site index, a measure of height growth rate has been used successfully in diameter increment models for older stand models (Hann and Larsen, 1991; Ritchie and Hann, 1985; Krumland and Wensel, 1981) and in younger stand models as well (Ritchie and Powers, 1993). Since site index is not a direct measure of productivity, it presents problems in young stands where previous site index estimates may not be available and estimation may be biased due to the influence of competing vegetation in the early stages of stand development (Newton and Hanson, 1998) or tree density (Scott et al., 1998).

Another productivity variable used is site productivity index. Huang and Titus (1994) defined site productivity index (SPI) as the height for a 20 centimeter reference point on a stem and used the variable in a diameter increment model (Huang and Titus, 1995). The 20 centimeter diameter reference height corresponds to the 50-year reference age for site index curves in Alberta (Huang and Titus, 1993).

In addition to site index and SPI, other authors have chosen to use physical site variables such as slope, aspect, elevation, and soil depth (Stage, 1976; Dolph, 1992b, 1988b). Stage (1976) used a simple trigonometric transformations on aspect and slope as well as indicators for habitat type to estimate the site index, showing that these endemic variables could replace site index as a measure of site productivity. The author failed to report fit statistics which weakens the argument. Dolph (1992b) simply included the endemic variables (slope and aspect) directly into the regression equations.

All other factors being equal, site productivity influences the rate of growth in forest trees. On poor sites growth rates are low and the inverse is true on height sites. Stands that occupy high sites will exhibit an earlier culmination of mean annual increment, experience density induced mortality sooner and will produce stands with taller dominant height curves at a given age. Estimates of site quality, often measured in site index, or a height at a given base age, are commonly used in diameter growth models (Hann and Larsen, 1991; Ritchie and Hann, 1985; Nystrom and Kexi, 1997; Krumland and Wensel, 1981).

1.3.2 Height Growth

It can be argued that height growth is the most influential variable for the early development of a stand. Height growth has historically focused on the development of site index or dominant-height-growth equations (Dolph, 1992a; Ritchie and Hann, 1990; Hann and Ritchie, 1988; Hann et al., 1987). While height growth in older stand models has been widely studied, height growth in young plantations has received little attention. Problems due to the influence of competing vegetation, precision problems in site index that have included young trees and the assumption that height growth is a simple polynomial curve have all contributed to the lack of sophisticated model development for height growth models in young trees.

Again, explanatory variables in height growth equations typically can be categorized to represent plant size, vigor, competitive position, population density, and site productivity (Wykoff, 1990; Vanclay, 1994). A summary of how various plant attributes have been used in height equations is presented in Table 1.2.

1.3.2.1 Plant Size

Simple field measurements are often used as measures of tree size. Total height (THT), stem diameters and their transformations are the most common (Ek and Monserud, 1974; Stage, 1975; Dolph, 1988a; Wensel et al., 1987; Ritchie and Hann, 1990; Opalach et al., 1990).

Age can also be used as a surrogate for plant size but does have disadvantages when compared to simple field measurements such as diameter and height. Determining age often requires extracting increment cores, particularly when stands are in the stem exclusion phase and self-pruning prevents obtaining whorl counts. In some cases, such as hardwoods, whorl counts are not possible as a result of the morphology of the plants. Stem diameter and heights are easier to measure, when compared to age, because the cost of extracting increment cores from older trees is more costly than accurate height measurements.

1.3.2.2 Vigor

Tree vigor attributes that describe a plant's ability to utilize resources, can be characterized by crown attributes, current growth rates, and previous growth rates. Crown ratio is often used to characterize tree vigor because of the inherent association with dominance and the influence on a potential function as a modifier (Ritchie and Hann, 1990, 1986; Wensel et al., 1987). Ritchie and Hann (1986) and Wensel et al. (1987) successfully utilized crown ratio as a measure of plant vigor for a height increment model where crown ratio was held constant for the growth period.

Other forms of vigor have been utilized with success such as past diameter and height growth rates. Dolph (1988a, 1992a) incorporated diameter growth and basal area increment (BAI) into height growth equations for mixed conifers and red fir in the Sierra Nevada Mountains and Northern California, respectively. Arney (1985) used diameter growth and total height growth as measures of tree vigor in an older forest growth model. Opalach et al. (1990) used relative height (RH) in a regeneration model for coastal Oregon forest plantations. Stage (1975) included diameter growth with crown ratio in height growth equations. Ritchie and Powers (1993) used height relative to other confiers (RH(ITC)) and height relative to competing vegetation (RVH) as a measure of both tree vigor and competitive position.

Other authors have included indicator or class variables to describe plant vigor in height growth models. In addition to common measures of tree vigor, Dolph (1992a) included indicator variables for the dwarf mistletoe rating (DMR) system developed by Hawksworth (1977). Ferguson and Adams (1980) found that a damage code was significant in a height growth model in advanced grand fir regeneration. Ek and Monserud (1974) added a measure of shade tolerance (STI) as an indicator of plant vigor under various levels of competition.

1.3.2.3 Competitive Position

Variables that describe the competitive position and the severity of competition are often called "one-sided" competition variables (Vanclay, 1994). Vanclay (1994) describes one-sided competition as the influence of larger plants on smaller plants, or plants within a certain zone of influence larger than the subject plant. Since this form of competition is primarily thought of as competition for light, variables that are related to crown dimensions are often used.

Crown ratio, in addition to being used as a measure of tree vigor, has also been used extensively in height growth equations to describe the competitive position in relation to other trees (Dolph, 1988a, 1992a; Hann and Ritchie, 1988; Ritchie and Hann, 1990; Wensel et al., 1987; Hann and Larsen, 1991). Crown ratio has many advantages over other measurements by including a measure of the competitive position and tree vigor in a single variable without introducing multicollinearity. Crown competition in trees larger and basal area (related to leaf area) in trees larger have been used as well to express one-sided competition.

Wensel et al. (1987) used crown ratio and the crown closure at 66 percent of subject tree's height (CC_{66}) to describe the competitive position of the tree. Hann and Ritchie (1988) found the equation developed by Wensel et al. (1987) under-predicted height growth rates for trees in the lower crown classes and used CCH, the crown closure at the tip of the subject tree, raised to various powers in addition to crown ratio as a measure of the competitive position with favorable results. Opalach et al. (1990) utilized the ratio of DBH over top height.

Crown competition factor (Krajicek et al., 1961) in larger trees (CCFL) has also been used to describe the competitive position of a tree in height increment models (Ritchie and Hann, 1985). Crown competition factor was originally developed to describe the level of above ground competition or density. Ritchie and Hann (1985) partitioned CCF into size classes to analyze the effects of crown competition from stems larger than the subject stem. The authors found that the best combination of variables was CCFL and stand basal area (related to CCF), which could mean that the best variable for describing the competitive position of a tree, when describing height increment, is CCFL.

Basal area in trees larger (BAL) is another measure to describe the competitive position of a tree in height growth models (Wykoff et al., 1982). Wykoff et al. (1982) found BAL, a surrogate for leaf area in larger trees, was significant in height growth equations for young trees.

1.3.2.4 Density

In addition to competition for light resources, measures of density, or "two-sided" competition, competition for resources below ground or with plants in the same size or competition class (Vanclay, 1994), also plays a role in describing height growth. That is to say, competition among cohorts as determined by density variables such as basal area and stems per unit area (TPA), including their transformations, have been successfully used in older tree models as a measures of stand density (Dolph, 1988a, 1992a; Wensel et al., 1987; Arney, 1985; Ek and Monserud, 1974). Other common measurements of stand or plot density include stand density index (SDI) (Reineke, 1933), relative density (RD) (Drew and Flewelling, 1979; Curtis, 1982), crown closure, crown competition factor (Krajicek et al., 1961), foliar density, and percent shrub cover.

As mentioned previously, crown competition factor (CCF) (Krajicek et al., 1961) has been used to describe density in height growth equations for trees (Hann and Ritchie, 1988; Arney, 1985; Ritchie and Powers, 1993; Wykoff et al., 1982; Wykoff, 1990). Also, measures of foliar weight have been analyzed in previous attempts to describe density. Ritchie and Hann (1985) used foliage weight to assess the competitive stress between trees but found that crown ratio was a better measure of the competitive position in describing diameter increment.

Reference	Size	Vigor	Competitive Position (One-Sided)	Density (Two-Sided)	Site Productivity
Dolph (1988a)	DBH THT	CR ΔDBH BAI	BAL $\frac{BAL}{DBH}$	SBA	Elevation Slope Aspect Site Index Latitude
Dolph (1992a)	THT	DMR CR ΔDBH	CR	SBA	Location Class Aspect Slope Elevation Site Index
Hann and Ritchie (1988)	THT	CR	CR CCH	CCF	Site Index
Ritchie and Hann (1986)	THT	CR	$\frac{H}{SH}$	N/A	Site Index
Arney (1985)	DBH Age	ΔDBH ΔTHT	SH	SBA CCF TPA	Site Index
Wensel et al. (1987)	DBH THT Age	CR	CR CC_{66}	PCTBA	Site Index
Opalach et al. (1990)	THT Age	RH	RH	PCTCOV	Site Index
Stage (1975)	DBH THT	CR ΔDBH	CR	N/A	Site Index Location Habitat Type
Ritchie and Powers (1993)	THT Age	RH(ITC)	ITC RVH	PCTSCOV CCF	Site Index
Ek and Monserud (1974)	THT CRAD	CI STI	CI STI	TPA	Site Index
Ritchie and Hann (1985)	DBH	CR ΔTHT	CR CCFL	SBA	Site Index
Hann and Larsen (1991)	DBH	CR	CR	SBA	Site Index
Wykoff et al. (1982)	THT	N/A	BAL	CCF	Habitat Type Location Slope Aspect
Brand (1986)	RH	RH	HP	PCTCOV FD	N/A
Wagner (1989)	RH	THT	RH CI	PCTCOV	Site Index

Table. 1.2: Common explanitory variables used in height increment models

In addition to individual tree measurements such as relative height (e.g. tree height to top height) and crown ratio, aggregate measures of density observations can be included in height growth equations. Percent cover has been widely used in height growth equations to describe the density of competing shrubs in single-tree-aggregate-shrub models (Ritchie and Powers, 1993). Percent cover has inherent attributes that make the measurement appealing. The value is bound between zero and some upper limit depending on the definition of percent cover, normally between 0 and 100, and can be used in conjunction with tree crown dimensions to describe the available area for crown expansion. Percent cover can then be further divided by plant form (shrub, hardwood, conifer) and species specific values. If percent cover is broken down into species specific values, a modifier describing the influence can be assigned for a relative height to describe the influence of different levels of competition and density by species for various relative heights (Wagner, 1989; Opalach et al., 1990). Wagner (1989) found that the best measure of the influence of competing vegetation on height growth was the percent cover of woody shrub species taller than the subject tree for Douglas-fir in plantations. In most cases, percent cover can be derived by computing sum of the crown area from the individual plant observations and dividing the value by the area the plot represents.

1.3.2.5 Site Productivity

Typically site productivity or quality is measured by site index. This has appeal in height growth models because of the implied nature of predicting height growth from age and site by differentiating the cumulative height growth function. Site index equations can also be expressed as age functions in determining the growth effective age (GEA) from height and site index for a stand (Ritchie and Hann, 1990). Often GEA is defined as the age at which a tree would be at a certain height for a given site index. The measure is analogous to the inverse of the site index-height function where age becomes the dependent variable.

Site index curves have been developed for many species found within the study area (Hann et al., 1987; Dolph, 1987, 1991; Powers and Oliver, 1978; Biging and Wensel, 1985). Site index functions have an advantage of measuring the maximum height growth rate for site trees. Also, the functions can be solved to estimate age from site and height observations, thus allowing the equations to be used in multi-aged stands in an age-independent manor using a GEA variable.

Other independent variables have been incorporated into measures of site productivity in height growth models as well. Dolph (1988a, 1992a) used measures of elevation, slope and aspect in addition to site index to predict height increment in young-growth mixed conifers and Red fir. Stage (1975), in addition to using site index, incorporated a parameter for the population and habitat type into equations developed for predicting height increment. Wykoff et al. (1982) used habitat, location, slope and aspect, without using site index to predict height growth in young trees. Uzoh (2001) used water holding capacity (WHC) as a measure of site productivity in a height increment models.

1.4 MODELS

Model form is essential in developing an effective and robust model that accurately describes the dependent variable across a wide range of values.

The model form can be linear or non-linear and may reflect forms that mimic the biological phenomenon, such as a peaking function for growth functions. Functions for a limited range in such a system, such as diameter increment functions for young stand models, may or may not fit these criteria, but a robust model will be able to handle predictions well past the limits of the model dataset. For example, a height growth model that has the form

$$\Delta THT = \beta_0 + \beta_1 THT + \beta_2 THT^2 + \epsilon \tag{1.4}$$

should contain a negative β_2 coefficient so that the function would "peak" at some value when height growth is greatest. This function would not extrapolate well, however. If the β_2 coefficient was negative the predicted value of ΔTHT would be negative for very tall trees and very short trees. If the β_2 coefficient is positive, the curve would decrease for small trees and then increase for larger and larger trees. Clearly this model is limited.

Whether the nature of the model is theoretical or empirical, it should be formulated to provide meaningful predictions across a wide range of explanatory variables. The explanatory variables should be carefully chosen from large datasets and the model forms should represent biologically meaningful relationships (Vanclay, 1994).

1.4.1 Empirical Equations

Linear regression equations make up the bulk of models used to describe both diameter and height increment. While empirical in nature, linear regression models can provide biologically meaningful behavior (Vanclay, 1994). An empirical model can be developed using the linear regression equation,

$$Y = b_0 + b_1 x_1 + b_n x_n + \ldots + b_n x_n + \epsilon \tag{1.5}$$

or in matrix notation,

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{1.6}$$

where **X** is an array of explanatory variables and **Y** is a vector of the response variable. The error term, ϵ is assumed to be $N(0,\sigma^2)$. The equation can be solved using linear algebra and the parameter estimates are the Best Linear Unbiased Estimates (BLUE) when the assumptions of Ordinary Least Squares (OLS) are met. The **X** and **Y** may contain transformed variables that are not linear, such as DBH^2 and $\ln(THT)$, but the model is said to be linear because it is linear in the parameters. The OLS models contain unique solutions and are robust even when the assumptions of OLS are not met (Vanclay, 1994).

1.4.2 Theoretical Equations

Equation (1.4) is normally considered a curvilinear relationship because the relationship of the response and explanatory variables is still a linear function (Ratkowsky, 1990). While it does attempt to demonstrate the explanation between height and height growth, equation (1.4) might have missing explanatory variables such as density, competitive position, and vigor. This is referred to as model specification error, or simply specification error (Kmenta, 1997; Greene, 2000). In contrast to empirical equations, theoretically based models attempt to explain, using a mathematical function, the relationship of a hypothesis to the explanatory variables (Vanclay, 1994).

Scaling laws (Enquist et al., 1998; West et al., 1997) and size-density relationships (Puettmann et al., 1993; Weller, 1989; Gorham, 1979) have been used to develop growth models based on the relationships among the available growing space, plant energetics, and plant size attributes with success. These theories have appeal when addressing mortality models but are outside the scope of young tree models where maximum size-density relationships and self-thinning do not occur.

Theoretical equations are appealing because they allow modellers to develop a framework instead of developing component models to explain a single characteristics of a complex system without regard for functional relationships. It is for these reasons that theoretical models are rarely used for the development of regeneration models.

1.5 MODEL FORMS

Regardless of the decision to develop an empirical or theoretically based model, the equation must fit into one of three forms: linear, intrinsically linear, or intrinsically nonlinear forms (Ratkowsky, 1990). Linear model forms, those forms that are expressed as equation (1.5) are considered linear in that no transformation is required on the response variable to obtain the best linear unbiased estimates (Ratkowsky, 1990). These are rare in forestry increment models due to the non-normal nature of the dependent variables, limited range of data and in many cases the curved nature of the relationships such as site index functions. The intrinsically linear model form, often a logarithmic transformation on the response variable, is commonly used in increment models (Wykoff, 1990; Dolph, 1992a, 1988b,a; Ritchie and Hann, 1985). Those models that cannot be expressed linearly and must be solved using the general form of least squares are known as nonlinear models. These models, often derived from theoretical relationships (see 1.4.2) are considered to aid in a better understanding of the underlying phenomenon regarding increment models (Pienaar and Turnbull, 1973). Few nonlinear models are native to forestry and are typically developed in other fields of study and applied to forestry. The datasets in these other fields of study are vast and usually contain many generations of growth data before a model is fit. In forestry, this is rarely the case. It is more difficult to obtain nonlinear parameter estimates than linear parameter estimates since standard regression packages are typically developed for linear models and it is more difficult to develop regression diagnostics in those packages without having access to the underlying algorithms.

1.5.1 Linear Models

Until the development of computer software that could perform nonlinear regression, the majority of models developed to predict diameter and height increment were linear models (Wensel et al., 1987; Dolph, 1988b,a, 1992a; Ritchie and Hann, 1985; Stage, 1975, 1973b).

Stage (1973b) used a linear model for both basal area and height increment for lodgepole pine to facilitate projection intervals beyond the sample interval. The basal area increment model, predicting the rate of increase in DBH, was a re-expression of the logarithm of basal area growth. The response variable was fit using a logarithmic transformation resulting in homoscedastistic errors. The linear height increment model used DBH, THT and diameter growth as the independent variables. The model uses a "location" variable as a method to calibrate the effects of differing regions.

The regression diagnostics and study behind linear regression models is vast to say the least. There are many tests for each of the assumptions, each with their respective strengths and shortcomings. In the next few sections, I present a brief description of the dangers of violating some of the assumptions and list a few tests and remedies.

1.5.1.1 Normality

Least squares methods assume errors are normally distributed with a constant variance $\epsilon = N(0, \sigma^2)$. Many of the statistical tests for the unbiased parameter estimates rely on this assumption. Non-normal error distributions can lead to incorrect confidence intervals and hypothesis test results (Neter et al., 1996; Draper and Smith, 1998). The normality tests can be divided into two classes: directions test and omnibus tests. Directions tests, which includes the tests developed by D'Agostino and Tietjen (1973), D'Agostino and Pearson (1973) and Geary's Test (Geary, 1935), assume prior knowledge about the departures from normality. The omnibus tests, which includes the Shapiro-Wilk Test (Shapiro and Wilk, 1965), the D'Agostino test (D'Agostino, 1971), and the Kolmogorov-Smirnov Test (Shapiro et al., 1968), do not assume the form of the departure from normality.

Departures from normality can be corrected for by transformations of the dependent variable, the development of estimators based on the underlying distribution, such as maximum likelihood, or by applying adjustments to the standard tests, such as the F-test, to account for the departures (Prentice, 1974).

1.5.1.2 Homogeneous Variance

If the residuals for the linear regression are not homogeneous, that is the dispersion of the residuals about the mean regression line is not constant across the range of predictor and predicted variables, the parameter estimates may be unbiased, but they are inefficient as will be the variance estimates of the parameter estimates.

The tests for heteroskedasticity can be classified into two forms. Constructive tests provide information about the form of the heteroskedasticity in addition to the presence of heteroskedasticity and non-constructive test only test for heteroskedasticity. The two more popular non-constructive tests tests used in the development of forest growth models includes the Levene Test (Levene, 1960) which is very robust to non-normality and the Goldfeld-Quandt Test (Goldfeld and Quandt, 1965) in which the form of the heterogeneity is precisely known. The methods to adjust for heterogeneity are numerous. Transformations using logarithms or weighted regression are common to reduce heteroskedastic errors about the mean regression line.

1.5.1.3 Independent Explanatory Variables

The consequences of developing linear regression models that contain errors in the independent variables can include biased and inefficient parameter estimates and predictions (Zellner, 1962). Methods for addressing stochastic independent variables include single equation techniques, as found in measurement error literature (Flewelling and Jong, 1994) and systems of equations literature (Tang et al., 2001; Hasenauer et al., 1998; Huang and Titus, 1999).

If at least two variables are perfectly correlated such that they are functions of each other, then there is a linear relationship between the variables and the resulting **(X'X)** matrix will be singular and the parameter estimates cannot be determined (Vanclay, 1994). In single equation models, an equation developed for a specific attribute such as tree volume may contain information that is also used to describe another attributes as well such as total height.

When "close to perfect multicollinearity" exists, even though the $\mathbf{X}'\mathbf{X}$ matrix has an inverse, many computer programs will fail to find the inverse due to the precision of the program. If the program does find an inverse, the diagonal elements of the inverse may be large as will the standard errors and confidence intervals and the t-ratios may be small, suggesting that few of the parameter estimates are significant from zero (Intriligator, 1978).

Some evidence of multicollinearity may be invalid signs based on prior knowledge, unusually large variance inflation factors (VIF), and obvious relationships among the variables in the correlation matrix (Draper and Smith, 1998). The most common remedy is to drop one of the variables that are correlated. For a more detailed examination and solution to the multicollinearity problem, Belsley's method may be used (Draper and Smith, 1998).

Explanatory variables are often correlated within a set of regression equations such as those found in growth and yield models. In systems of equations, multicollinearity has a special significance because of the relationships among the variables and error terms in the set of regression equations. If unaccounted for, these relationships will produce biased and possibly inefficient parameter estimates.

1.5.1.4 Serial Correlation

If serial correlation is present among the residuals, OLS will produce unbiased and consistent parameter estimates but not efficient parameter estimates nor variance estimates (Kmenta, 1997). Durbin and Watson (1950) developed a test for serial correlation that assumes first-order collation. Durbin (1970) developed an exact test for serial correlation although it was determined to be less powerful than the test developed by Durbin and Watson (1950).

To correct for serial correlation, Generalized Least Squares (GLS) can be applied if the the variance-covariance matrix of the regression is available.

This often not the case and other methods can be applied such as Maximum-Likelihood Estimation (MLE) (Kmenta, 1997).

1.5.2 Nonlinear Models

Nonlinear model forms, those model forms that cannot be linearized by transformations, are typically developed from theoretical relationships and may produce better predictions when extrapolated beyond the original range of data (Vanclay, 1994). Nonlinear models are potentially more difficult to fit and, as is the case with linear models, when fit to data that cover a narrow range, these models may not exprapolate well. Ratkowsky (1990) described the difficulties of fitting nonlinear models. The least squares estimates cannot be determined from explicit mathematical relationships. The surface of the sums-of-squares values for the parameter estimates may be highly curved for some model forms that are not close to linear.

These models do not necessarily produce a unique best unbiased solution (Vanclay, 1994) but can be re-parameterized so that the model behaves in a "close to linear fashion" (Ratkowsky, 1990). These close-to-linear forms can be fit using the Guass-Newton method to obtain the smallest sums-of-squares value resulting in parameter estimates that are unbiased, the errors are normally distributed and the variance estimates are lowest for the data (Ratkowsky, 1990).

The linearization method may never converge and parameter estimates my oscillate about some value. The steepest decent method may be very slow to converge after rapid initial progress, and the method is sensitive to the scale of the parameter estimates (Draper and Smith, 1998). Good starting values are required in a steepest-descent method (Draper and Smith, 1998) which may be problematic in new areas of development in forest models.

Often, these starting values are obtained from previously reported values in similarly published models. Another method, known as Marquardt's method, is often used to obtain parameter estimates by using a combination of the linearization and and steepest descent methods (Draper and Smith, 1998).

There are diagnostics for nonlinear models that can be used to assess the ease with which the model can be fitted. Bates and Watts (1988) developed curvature measures for nonlinearity which suggest how well linear approximation can estimate the nonlinear function. Box (1971) developed a formula for estimating the bias for estimated nonlinear regression coefficients. Hougaard (1985) developed an estimate of the skewness of the sampling distributions of the estimated regression coefficients. Many of the linear tests are applicable to non-linear fitting methods with large sample sizes.

Large sample sizes are required in order to obtain unbiased parameter estimates and normally distributed residuals and to make the same inferences about the parameter estimates as is the case in linear regression (Neter et al., 1996).

As computer programs have become more powerful, more nonlinear forest models have been developed in recent years (Hann et al., 1993; Wensel et al., 1987). These models are based on theoretical relationships, they extrapolate well and provide a framework for the further development of related forest models. As more data become available, these types of models will become more commonplace.

1.5.3 Systems of Equations

Tree attributes are rarely measured independently and because static and increment models are built at the same time as a set of equations and from the same data set, problems may be amplified. These problems may or may not be detected in single equation regression diagnostics. When multiple models are developed from the same database or are developed to work together, the error terms and variables may be related. If there is a relationship among the equations of a system the regression equations are said to be contemporaneously correlated (Kmenta, 1997; Greene, 2000).

In addition, when predicted variables from one equation are used as independent variables in another equation, the set of equations may be subject to simultaneity bias (Kmenta, 1997; Greene, 2000). Simultaneity bias is equivalent to the "errors-in-variables" problem described in Section 1.5. When the system of equations contains cross equation dependent variables, using OLS will produce inconsistent and biased parameter estimates. For example, consider the following two-equation system:

$$\begin{aligned} y_1 &= a_1 + b_1 y_2 + c_1 x_1 + \epsilon_1 \\ y_2 &= a_2 + b_2 y_1 + c_2 x_2 + \epsilon_2 \end{aligned}$$

In the second equation, y_2 is a dependent, or endogenous variable. Also, y_2 is a function of y_1 and ϵ_1 thus making both y_1 and y_2 jointly dependent. This set of regression equations is known as simultaneous and is subject to simultaneity bias due to the relationships among the dependent variables.

Sets or systems of equations can be divided into two types. Sets of equations that are related through the residuals are known as Seemingly Unrelated Regressions (SUR) (Zellner, 1962) because the test to determine if a relationship exists determines if there is in fact a relationship, or if they are "seemingly" unrelated (Kmenta, 1997). When equations are related by the variables used within the set of equations and thus, subject to the "errors-in-variables" problem, then the set of equations is said to be a system of simultaneous equations (Kmenta, 1997) and can be solved using an instrument-variables technique known as Two-Stage Least Squares (2SLS).

Estimation methods for systems of equations are similar to those for single equations systems. Parameter estimates for single and simultaneous equations may be obtained using Generalized Least Squares (GLS). When the regression equations are correlated through the error terms and there are no dependent regressors, then SUR can be used to obtain unbiased and consistent parameter estimates. If the regression equations are related through the residuals and the system contains dependent regressors, then 2SLS can be combined with SUR which is referred to as Three-Stage Least Squares (3SLS) (Zellner and Theil, 1962).

Further discussion and details regarding the methods for SUR, 2SLS, and 3SLS will be presented in the next chapter.

1.6 SUMMARY

Developing a single tree diameter and height increment model is no trivial task. In addition to the wide selection of explanatory variables available, model forms and estimation methods there are plenty of diagnostic tools to aid the development of increment models.

Many authors have developed adequate models based on a few simple explanatory variables as described in Sections 1.3.1 and 1.3.2. The most commonly used explanatory variables for developing these increment models are related to the variables themselves. For example, in diameter increment models, DBH, SBA, BAL, and relative diameter are used as explanatory variables. In height growth models, THT, CCFL, CR and CCH are commonly used. In young tree models some of the more complicated or derived variables are seldom significant but are included regardless to ensure robustness and well behaved extrapolation and may address model specification issues as well.

Both linear and nonlinear models have been developed and widely used for both diameter and height increment. While linear models have an inherent simplicity and the regression diagnostics are well understood, nonlinear models are gaining popularity in diameter and height increment models. Nonlinear models have the added appeal of mimicking the underlying theoretical behavior of the relationships and being more biologically realistic.

The estimation methods for obtaining unbiased, consistent, and efficient parameter estimates are readily available as well for both linear and nonlinear models. Again, nonlinear model diagnostics are less developed and many regression packages have yet to adopt them as part of standard fitting routines. And until recently, both linear and nonlinear sets of forest growth and yield models have addressed cross equation relationships or endogenous dependent variables correctly.

The next chapter focuses on the regression methods for systems of equations and these methods are used to develop a model for DBH and total height increment in Chapter 3.

2. SIMULTANEOUS EQUATION ESTIMATION METHODS FOR FOREST GROWTH MODELS

2.1 ABSTRACT

This paper gives a brief presentation of forest growth and yield systems of equations estimation and methods to obtain the best linear unbiased estimates (BLUE). The methods of Ordinary Least Squares (OLS), Seemingly Unrelated Regressions(SUR), Two-Stage Least Squares (2SLS) and Three-Stage Least Squares (3SLS) are presented as options for fitting a set of equations to a data set in a forest growth model context.

2.2 INTRODUCTION

Sets of regression equations, in the form of height growth (Ritchie and Hann, 1990; Hann and Ritchie, 1988), diameter growth (Stage, 1973b; Dolph, 1992b; Hann and Larsen, 1991), crown recession (Arney, 1985; Ritchie and Powers, 1993), mortality (Krumland and Wensel, 1981) and perhaps ingrowth, are often used to predict the growth of forest trees and stands. Typically, these equations are developed individually and applied as a system of equations. Traditional fitting techniques used to produce parameter and variance estimates for these models may produce inefficient or inconsistent parameter estimates as a result of the relationships among the variables and how relationships are used within the the system of equations. The resulting parameter estimates, variance estimates and predictions may either be biased, inconsistent, or both (Furnival and Wilson, 1971; Lemay, 1990; Greene, 2000; Kmenta, 1997).

Methods to obtain unbiased and consistent parameter and variance estimates have been developed and used in forestry applications (Furnival and Wilson, 1971; Lemay, 1990; Borders and Bailey, 1986). This chapter will present and define the key concepts and terms, give a brief overview of the fitting techniques used to obtain unbiased and consistent parameter and variance estimates, and offer an example specific to growth prediction in forestry.

2.2.1 Stand Growth Example

To initiate an introduction to systems of equations, a presentation to illustrate the issues is required. Furnival and Wilson (1971) presented the following yield model,

$$\ln(H) = \beta_{11} + \beta_{12}\ln(A) + \epsilon_1 \quad (2.1)$$

$$\ln(N) = \beta_{21} + \beta_{22}\ln(D) + \epsilon_2 \quad (2.2)$$

$$\ln(B) = \beta_{31} + \beta_{32}\ln(H) + \epsilon_3 \quad (2.3)$$

$$\ln(F) = \beta_{41} + \beta_{42}\ln(D) + \epsilon_4 \quad (2.4)$$

$$\ln(B) = \ln(K) + \ln(N) + 2\ln(D) \quad (2.5)$$

$$\ln(V) = \ln(F) + \ln(H) + \ln(B) \quad (2.6)$$

where

V = volume, in cubic feet

H = average stand height, in feet

A = average stand age at breast height

B = basal area, in square feet per acre

D = diameter of the tree of average basal area

N = number of trees per acre

F = cylindrical form factor

$K = \frac{\pi}{576.0}$

$\ln(x)$ = natural logarithm of x

Equation 2.1 was chosen by the authors by trial and error (Furnival and Wilson, 1971) and Equation 2.2 is the widely used equation for the relationship between size and maximum density when applied to forestry (Reineke, 1933). The remainder of the equations are from various sources (Schumacher and Coile, 1960). The last two identities provide a relationship for trees per acre, average diameter, mean stand form class, and stand height.

A cursory examination reveals that many of the equations contain variables, predicted in one equation and used as a predictor in another. But what determines which variable is used in which equation; observed or predicted? For example, say we fit equation 2.1 and obtain predicted values for height. Do we use the predicted height or the observed height when predicting the basal area in equation 2.3? If we use the observed values of height in equation 2.3 and the parameter estimates for equation 2.1 are biased, then we should expect that our parameter estimates for equation 2.3 to be biased as well. If we used the predicted values for height in equation 2.3, assuming that the model for equation 2.1 contains the best linear unbiased estimates (BLUE) then we could expect that the estimates for the equation 2.3 to be BLUE as well.

Using a biased estimator in one equation and then as a predictor in another equation introduces bias across the equations and is known as simultaneity bias. The origins of simultaneity bias are found in the "errors-in-variables" problem as described by Kmenta (1997). When the assumption that the independent variables are measured without error is violated, the parameter estimates are biased, inconsistent, or both. At the very least, using Ordinary Least Squares (OLS) on individual equations is inefficient when describing a related system of equations.

To remove the potential simultaneity bias that may arise from using the observed height in equation 2.3, the fitted values from equation 2.1 are used as the observations when fitting equation 2.3 (Furnival and Wilson, 1971). Since the parameter estimates for equation 2.1 are BLUE, the predicted values for height in equation 2.3 are assumed to be measured without error, as there is no bias in the "observations", thus eliminating simultaneity bias. The predicted values are assumed to be measured without error because the expected value for the predicted height values is the mean.

2.2.2 Individual Tree Example

Simultaneity bias is not unique to stand level models nor models that are linear in the parameters and variables (Parks, 1967). Any system of equations is subject to simultaneity bias depending on the relationships of the variables. Arney (1985) developed an individual tree or diameter class nonlinear set of equations,

$$H = 1.37 + B_1 e^{B_2 DBH^{B_3}} \tag{2.7}$$

$$\Delta H = \Delta TOP \left(\frac{H}{TOP}\right)^{B_1} \left[1.0 - \left(\frac{CCF}{B_2}\right)^{B_3}\right] \tag{2.8}$$

$$\frac{\Delta DBH}{\Delta TOP} = B_1 \left(\frac{CCF}{100}\right)^{B_2} \left(1.0 - e^{B_3 \frac{DBH}{TOP}}\right)^{B_4} \tag{2.9}$$

$$TPA = TPA_{100}(0.78 + 0.22(1.29 - 0.29CCF)^{0.25}) \tag{2.10}$$

$$CCF = \frac{100\Sigma_n^i GS_i}{A10000} \tag{2.11}$$

$$GS = \frac{\pi}{4} CW^2 \tag{2.12}$$

$$CW = (1.19 + 24.7(1.0 - e^{-0.001DBH})) \tag{2.13}$$

where

DBH = diameter at breast height

H = total height of the subject tree or DBH class

TOP = top height of the stand

CW = crown width of the subject tree or DBH class

CCF = crown competition factor

TPA = trees/acre

A = plot area

GS = growing space required for the subject tree or DBH class

and ΔTOP is the height growth of the site trees derived from a site index equation (King, 1966; Arney, 1985).

As was described in the previous example, errors are introduced into the independent variables (the "errors-in-variables" problem) causing parameter and variance estimates to become bias and potentially inconsistent. If the disturbances (or residuals) in the equations are related (known as contemporaneous correlation), then the parameter estimates may be asymptotically inefficient (Kmenta, 1997; Greene, 2000). The classical regression assumption that disturbances are unrelated may not hold in the case where multiple regression equations are developed to describe different but related components of a forest system.

This set of relationships also contains many nonlinear models which can increase the likelihood of producing inconsistent and inefficient parameter estimates due to the relationship among the variables and the disturbances.

2.2.3 Other Examples

The technique of using predicted variables as predictors is common. This is often done because the cost of measuring a full set of variables for each tree are economically prohibitive.

Predicted diameter growth has been used as an independent variable in height growth models for tree species in Idaho (Stage, 1975, 1973b). Computed or predicted ages, instead of measured stem age, have been used in sets of growth models as well (Ritchie and Hann, 1986; Hann and Ritchie, 1988). Crown width, which is typically used to compute stand level crown attributes is commonly predicted from DBH observations (Bella, 1971; Martin and Ek, 1984)

Bella developed a set of competition models that included DBH as a predictor and then used the amount of competitive influence as a predictor in a diameter increment model (Bella, 1971). The competitive influence was a function of DBH as well. The models could be used to estimate both diameter increment and inter-tree competition at the same time, but where fit independently.

Martin and Ek (1984) developed a height and diameter growth model to describe the behavior of red pine stands for silvicultural analysis. Site and age were used in both the diameter and height growth equations. The authors did not examine the potential relationships between the two equations.

These cross-equation relationships are commonly found in sets of equations that are developed for single-tree or stand growth models such as inter-tree competition and potential growth rates (Krumland and Wensel, 1981; Opalach et al., 1990; Burkhart et al., 1987; Ek and Monserud, 1974). The relationships occur mostly due to the inclusion of CCF or CCH in the height growth functions where CCF and/or CCH is calculated from DBH. When equations are developed for a single static attribute, such as crown dimensions (Paine and Hann, 1982; Gill et al., 2000; Farr et al., 1989), stem taper (Wensel and Olsen, 1993), or height-diameter (Larsen and Hann, 1987; Curtis, 1967), these concerns are minimal.

The next section will present the basic definitions and concepts for fitting groups of equations as a set or system of equations.

2.3 DEFINITIONS

To develop an understanding of the issues related to fitting systems of equations, some basic definitions and concepts must be presented to the reader. This section presents basic definitions and issues related to systems of regression equations.

2.3.1 Basic Equation Relationships

A system of equations is said to be simultaneous if there are variables that appear on the left hand side of one equation and again on the right hand side of another equation (Lemay, 1990). Fitting techniques to address simultaneity bias for both linear (Zellner and Theil, 1962) and nonlinear (Kelejian, 1971; Gallant, 1975) models have been developed. The first method to address simultaneity bias Two-Stage Least Squares (2SLS) which is a special case of an instrumental variables (IV) method (Zellner, 1962).

If the error terms in a set of regression equations are related, that is to say, the variance-covariance matrix for a set of equations is not a diagonal matrix, then the set of regression equations are said to be correlated through the error terms, or contemporaneously correlated (Lemay, 1990). Methods to obtain parameter estimates when a system is contemporaneously correlated are available (Zellner, 1962; Parks, 1967). The most common method used to address contemporaneous correlation, developed by Zellner (1962) is known as Seemingly Unrelated Regression (SUR).

When a set of regression equations contains related error structures (contemporaneous correlation) and variables that occur on both the left and right hand side (simultaneity bias), a combination of 2SLS and SUR, or Three-Stage Least Squares (3SLS), can be used to obtain unbiased, consistent and efficient estimates.

As there are techniques for solving a single regression equation where the disturbances contain serial correlation and heterogeneous variance using Generalized Least Squares (GLS), there are similar methods for systems of regression equations using Multi-Stage Least Squares (MSLS) (Lemay, 1990; Amemiya, 1977), Full-Information Maximum Likelihood (Parke, 1982).

Applications of the methods described above have been adopted to specific problems in forest growth models for stand level models (Furnival and Wilson, 1971; Borders, 1989; Borders and Bailey, 1986; Daniels and Burkhart, 1988) and single tree models (Lemay, 1990; Hasenauer et al., 1998; Huang and Titus, 1999; Rose and Lynch, 2001).

The method used to obtain parameter estimates is dictated by how the variables are related within the system of equations and the relationship of the equations. The next section will define how variables are classified to determine the various relationships among the equations in a group of equations. A group of equations that are related will be refered to as a set of equations and and a set of equations that contains endogenous variables as explanitory variables as a system of equations.

2.3.2 Variable Types

There are two main classes of variables in a system of equations. Variables that are determined outside a system are referred to as predetermined variables. In this case "outside" means a value that was not computed from another equation within a set of equations. In contrast, endogenous variables are those variables that are determined within the system (Kmenta, 1997; Greene, 2000). Predetermined variables can again be classified into lagged endogenous and exogenous, or variables that were determined during a previous prediction and those variables that are truly determined outside the system, respectively.

Since many forest models contain endogenous variables, examples abound. Growth rates, crown dimensions, biomass, heights from height-diameter relationships are a few examples. For this presentation, any variable that is estimated, not measured, is considered endogenous.

Since predetermined variables can be broken into exogenous and lagged endogenous, it helps to think of the forest stand in the context of ecological systems. Incoming solar radiation would be an example of an exogenous variable, as would precipitation in most cases. Regardless, there are few truly exogenous variables in forest ecosystems.

Lagged endogenous variables, those variables that were predicted from a previous estimation are more common when referring to sets of equations used to predict future stand conditions over many periods. For example,

$$DBH_{t+1} = f(DBH_t) \tag{2.14}$$

where $f(DBH_t)$ is some function. After the first iteration, DBH has become a lagged endogenous variable because the original value of DBH is no longer used to predict the future value of DBH. In fact, with regard to forest simulation, most variables that are considered exogenous are actually lagged endogenous. Current DBH is often considered an exogenous variable since modelers often make the assumption that growth is based on tree size (Hann and Larsen, 1991) when in fact tree size is the sum of all previous tree growth increments. So, in application, the current diameter is the sum of the all the previous diameter growth estimates plus the initial measured diameter.

The concept of a lagged endogenous variable is analogous to a spreadsheet program that uses the previous cell in a column to compute the value of the next cell. For example, when filling in a column of values to chart, the user places a zero in the A1 cell. Next, the user enters the formula "=A1+1" and then copies the formula into the 20 cells below. When the user looks at the formula in any cell, the result is similar to "=the value of this cell is the value of the cell above plus one".

2.3.3 Model Types and Model Forms

When describing a set or system of regression equations, the various fitting methods require the equations to be presented in either the structural or reduced form (Kmenta, 1997). An equation within a system that contains variables that were predicted from another equation, thus exhibiting the errors-in-variables, is called the structural form (Kmenta, 1997; Lemay, 1990). When the independent variables are not stochastic, that is to say, there is no error in the independent variables, the model form is called the reduced form (Kmenta, 1997).

These concepts are important for two reasons. First, the form conveys information about the system of equations. Investigators are interested in the parameter estimates that are meaningful to the variable being investigated. For example, when looking at stand size-density relationships, variables that are commonly used are average stand diameter ($\bar{D}$), diameter at breast height (DBH) and stand density, N, measured in trees per unit area. These unadulterated values are meaningful to foresters as they are values in the literature and are used everyday. In another form, say the reduced parameter estimates, may look like $\ln\left(\frac{N+\epsilon_t}{DBH}\right)$, which is more difficult to interpret. Second, and more importantly, the form aids in determining if and how the system can be solved. The reduced form equations are typically used to determine the identification of a system of equations.

While reduced form equations are more prevalent in econometrics texts, there are virtually no cases in which reduced equations are found in forestry

(Borders and Bailey, 1986). Borders and Bailey (1986) suggest that forestry relationships in the form of the structural equations contain certain *a priori* restrictions that demonstrate the structural relationships are an accurate representation of the actual relationships. Thus, since the structural equations contain all the required information to mimic the theoretical relationships, the requirement to obtain the structural parameter estimates from the reduced form equations is no longer necessary. In addition, current statistical software packages do not require reduced form equations and the user can enter into a list which variables are endogenous and exogenous. Also, forestry datasets usually contain far more than a handfull of observations which negates the requirements for testing for rank and order conditions for a system of equations, in most cases for forestry datasets. For an in depth review of the minutiae of reduced form equations and rank and order conditions, see Kmenta (1997) and Greene (2000). For this study, each equation contains at least equal or more excluded exogenous variables as included endogenous variables, thus meeting the order condition for each of the equations which is a necessary condition for identification (Greene, 2000).

2.4 ESTIMATION TECHNIQUES

Estimation techniques for systems of equations can broken down into two categories. Those sets of equations that can be estimated without regard for the relationships between the disturbances are known as single equation methods (Kmenta, 1997; Greene, 2000). Single equation methods produce parameter and variance estimates one equation at a time or independently. For single equation estimation, common choices include Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS) (Greene, 2000; Kmenta, 1997).

Alternatively, those methods that address the relationships of the disturbances and thus estimate all of the equations in a system simultaneously are known as simultaneous methods (Kmenta, 1997; Greene, 2000). Common methods are Seemingly Unrelated Regressions (SUR), Three-Stage Least Squares (3SLS), and various adaptations of these two (Greene, 2000; Kmenta, 1997; Lemay, 1990).

Other methods such as Limited Information Maximum Likelihood (LIML) for single equation estimation and Full Information Maximum Likelihood (FIML) for simultaneous systems are more complex, and are typically used to test other methods for their ability to meet the assumptions of GLS. These methods will not be presented here. Refer to Kmenta (1997) and Greene (2000) for detailed discussions on LIML and FIML.

2.4.1 Single Equation Estimation Techniques

Fitting techniques to address sets of equations that contain endogenous variables, and thus simultaneity bias, include Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS) (Greene, 2000; Kmenta, 1997). These techniques address the problems associated with measurement errors ("errors-in-variables") in independent variables and produce consistent estimates. The methods may not produce asymptotically efficient estimates due to the correlation between the equations that contain endogenous variables (Kmenta, 1997; Greene, 2000). These methods are typically defined as single equation estimation methods because the solutions can be obtained for the system by examining each equation independently and do not require information about the other equations in the system.

2.4.1.1 Ordinary Least Squares (OLS)

Ordinary Least Squares, while not directly addressing simultaneity bias, is used in the first stage of 2SLS to estimate the predetermined variables for each equation. Ordinary Least Squares will also produce consistent, efficient and unbiased parameter and variance estimates if the system of equations is recursive. An example of a recursve system could be one that contains two equations,

$$\widehat{CrownRatio} = \alpha_0 + \alpha_1 * Height$$
$$\widehat{Height} = \beta_0 + \beta_1 * Age + \beta_2 * Site$$

where the predicted height value is entered into the crown ratio equation and there is no relationship between the error terms of the regression equations. Since the predicted height values are assumed to be unbaised then the assumption is met that the heights (the predicted heights in this case) are measured without error in the crown ratio equation.

As a single regression equation can be estimated using

$$\mathrm{Y} = \mathrm{X}\beta + \epsilon \tag{2.15}$$

with a solution

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{2.16}$$

where $\hat{\beta}_{OLS}$ is a vector of regression estimates, $\mathbf{X}$ is a matrix of independent variables, $\mathbf{Y}$ is a column vector of dependent variable observations and ϵ some disturbance about the mean. Using matrix notation, it is possible to generate or "stack" M regressions together that can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix}, \tag{2.17}$$

or more compactly,

$$\mathrm{Y} = \mathbf{X}\beta + \epsilon \tag{2.18}$$

and the estimators are,

$$\hat{\beta}_{OLS} = (\mathbf{X}'(\hat{\mathbf{\Omega}}^{-1} \bigotimes \mathbf{I})\mathbf{X})^{-1}(\mathbf{X}'(\hat{\mathbf{\Omega}}^{-1} \bigotimes \mathbf{I})\mathbf{Y}) \tag{2.19}$$

and

$$\text{Asympt. Var-Cov}(\hat{\beta}_{OLS}) = (\mathbf{X}'\hat{(}\mathbf{\Omega}^{-1} \bigotimes \mathbf{I})\mathbf{X})^{-1} \tag{2.20}$$

where

$$\mathbf{\Omega} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \dots & \hat{\sigma}_{1M} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \dots & \hat{\sigma}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{M1} & \hat{\sigma}_{M2} & \dots & \hat{\sigma}_{MM} \end{bmatrix} \tag{2.21}$$

and

$$\hat{\sigma}_{ij} = \begin{cases} 0 & i \neq j \\ \frac{\epsilon_i' \epsilon_j}{n-p} & \text{otherwise} \end{cases} \tag{2.22}$$

where n is the number of observations for equation i, p is the number of independent variables, M is the total number of equations in the set of equations and $\otimes$ is the Kronecker product and **I** is an $\mathbf{M} \times \mathbf{M}$ identity matrix .

While OLS does not produce consistent and asymptotically efficient parameter estimates for a system of equations where simultaneity bias or contemporaneous correlation are present, OLS produces consistent and asymptotically efficient estimates if the system contains neither (Kmenta, 1997) or is a recursive system. An example might include a recursive set of equations such as crown ratio and height example previously described.

2.4.1.2 Two-Stage Least Squares (2SLS)

The Two-Stage Least Squares (2SLS) method uses, as the instruments for Y_j, the predicted values of Y_j on *all* the the *x's* (Kmenta, 1997; Greene, 2000). In 2SLS, the first step is to estimate the endogenous values and substitute the observations in the second stage to remove the simultaneity bias.

The application of 2SLS leads to unbiased parameter estimates, but does not produce consistent or efficient estimates if the disturbances of the various equations are correlated (Kmenta, 1997).

Using matrix notation, 2SLS estimators for equation j of a system may be expressed as

$$\hat{\beta}_{j,2SLS} = \left[\left(\mathbf{X_j'Z(Z'Z)^{-1}X_j}\right)\right]\mathbf{X'Z(Z'Z)^{-1}Z'}y_j \tag{2.23}$$

where $\mathbf{Z}$ is a matrix of all the instrumental variables, $\mathbf{X}$ is a matrix of the independent variables, y_j is a column vector of the dependent variable observations for equation j. As is the case in OLS, the asymptotic variance-covariance matrix for equation j, is

$$\hat{\sigma}_{ij} = \begin{cases} 0 & i \neq j \\ \frac{\epsilon_i'\epsilon_j}{n-p} & \text{otherwise} \end{cases} \tag{2.24}$$

2.4.2 Simultaneous Equations Estimation Techniques

Simultaneous estimation methods examine all of the structural equations while taking into account the relationships of the disturbances among the equations (Kmenta, 1997; Greene, 2000). The single equations methods do not use all the possible information about the system of equations. By ignoring the relationships among the various disturbances in the system, the parameter estimates may be subject to contemporaneous correlation. These types of errors can lead to inefficient and inconsistent variance and parameter estimates (Kmenta, 1997; Greene, 2000).

The SUR method is used to address contemporaneous correlation. If the equation system is simultaneous and there are endogenous variables, 2SLS and SUR can be combined to address both simultaneous equation bias (2SLS) and cross-equation correlation of the errors (SUR). This is called Three-Stage Least Squares (3SLS) (Zellner and Theil, 1962).

2.4.2.1 Seemingly Unrelated Regression (SUR)

The Seemingly Unrelated Regression (SUR) method links the disturbances of the equations by using GLS (Zellner, 1962) to include the cross equation variance-covariance information. The equations are stacked or estimated simultaneously like equation (2.17), hence the estimators for the parameters and the variance-covariance matrix are,

$$\hat{\beta}_{SUR} = (\mathbf{X}'(\hat{\mathbf{\Omega}}^{-1} \bigotimes \mathbf{I})\mathbf{X})^{-1}(\mathbf{X}'(\hat{\mathbf{\Omega}}^{-1} \bigotimes \mathbf{I})\mathbf{Y}) \tag{2.25}$$

and

$$\text{Asympt. Var-Cov}(\hat{\beta}_{SUR}) = \left(\mathbf{X}'(\hat{\mathbf{\Omega}}^{-1} \bigotimes \mathbf{I})\mathbf{X}\right)^{-1} \tag{2.26}$$

where

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \dots & \hat{\sigma}_{1M} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \dots & \hat{\sigma}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{M1} & \hat{\sigma}_{M2} & \dots & \hat{\sigma}_{MM} \end{bmatrix} \tag{2.27}$$

and

$$\hat{\sigma}_{ij} = \frac{\epsilon_i' \epsilon_j}{[(N - K_i)(N - K_j)]^{1/2}} \tag{2.28}$$

where m = 1,2,...,M. In this case, N is the number of total observations and K is the number of independent variables in equation m.

The SUR estimators are equivalent to the generalized least squares estimator (Kmenta, 1997) assuming the disturbances are normally distributed and non-autoregressive (Lemay, 1990). If the equations are in fact related, thus the name "Seemingly Unrelated Regressions", then SUR can be used to estimate parameters and the variance-covariance matrix more efficiently than by using OLS for each equation. If the equations are not related by means of the cross-equation variance-covariance matrix, in this case a singly banded diagonal matrix, then the residuals of the equations in the system are in fact not related, thus the equations are not related.

As with GLS, SUR can be applied to systems of equations that contain autoregressive and heteroskedastic error distributions in any of the equations (Parks, 1967).

Since using this approach involves more work than OLS and the payoff, in terms of efficiency and consistency may prove to be minimal, it would be nice to be able to estimate the amount of efficiency gained by utilizing SUR. A few authors have addressed this question. Zellner (1962) examined the effects of the related error terms to gauge the influence of using SUR when $\sigma_{ij} = 0$ for $i \neq j$ and found that there is no advantage to using GLS over OLS. In fact, in this case, the OLS estimates are equivalent to the SUR esimates. If the error terms are actually related, the more correlated the disturbances, the greater the efficiency (Greene, 2000; Zellner, 1962).

2.4.2.2 Three-Stage Least Squares (3SLS)

Three-stage least squares uses both SUR (to correct contemporaneous correlation) and 2SLS (to correct simultaneity bias) techniques to estimate the structural coefficients of a system of equations (Zellner and Theil, 1962). The first two stages of the 3SLS procedure are the same as 2SLS. The variance-covariance matrix is estimated from 2SLS residuals and the entire system is fit using SUR. The method produces consistent and efficient estimates of the variance-covariance matrix for the system of equations (Kmenta, 1997; Greene, 2000). It should be noted that when the disturbances among the equations are not correlated, then 3SLS is equivalent to two-stage least squares estimates (Kmenta, 1997; Greene, 2000).

Recalling from equation (2.23) ,

$$\hat{\beta}_{j,2SLS} = \left[\ (\mathbf{Z_j'X_j})(\mathbf{X_j'X_j})^{-1}(\mathbf{X_j'Z_j}) \ \right]^{-1} (\mathbf{X_j'X_j})(\mathbf{X_j'X_j})^{-1}\mathbf{X_j'}y_j \tag{2.29}$$

the 2SLS estimators can be "stacked", similar to the SUR method, to yield,

$$\hat{\mathbf{Z}} = \begin{bmatrix} \mathbf{X_1}(\mathbf{X_1'X_1})^{-1}\mathbf{X_1'Z_1} & 0 & 0 & 0 \\ 0 & \mathbf{X_2}(\mathbf{X_2'X_2})^{-1}\mathbf{X_2'Z_2} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X_M}(\mathbf{X_M'X_M})^{-1}\mathbf{X_M'Z_M} \end{bmatrix}$$

$$\hat{\mathbf{Z}} = \begin{bmatrix} \hat{\mathbf{Z}}_1 & 0 & 0 & 0 \\ 0 & \hat{\mathbf{Z}}_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\mathbf{Z}}_\mathbf{M} \end{bmatrix}$$

and using SUR to perform the final estimates to correct for contemporaneous correlation yields,

$$\hat{\beta}_{3SLS} = \left[\hat{\mathbf{Z}}'\left(\mathbf{\Sigma}^{-1}\bigotimes\mathbf{I}\right)\hat{\mathbf{Z}}\right]^{-1}\hat{\mathbf{Z}}'\left(\mathbf{\Sigma}^{-1}\bigotimes\mathbf{I}\right)\mathbf{Y} \tag{2.30}$$

The asymptotic variance-covariance matrix for the estimator is

$$\text{Asympt. Var-Cov}\left[\hat{\beta}_{3SLS}\right] = \left[\hat{\mathbf{Z}}'\left(\mathbf{\Sigma}^{-1}\bigotimes\mathbf{I}\right)\hat{\mathbf{Z}}\right]^{-1} \tag{2.31}$$

which would be estimated from the inverse matrix in equation 2.30.

For normally distributed disturbances, 3SLS produces asymptotically efficient estimators as does the FIML estimators (Greene, 2000).

2.5 A FORESTRY EXAMPLE

A simple example to demonstrate how to obtain the parameter estimates for a system of regression equations is in order. Let us examine the relationship of a height growth and diameter growth model for young plantations. As we have seen from Chapter 1, there are plenty of variables that can be used to describe height growth and diameter growth. Often height growth is used in diameter growth equations. This is an example of such a model. Lets assume a simple peaking function for both height growth and diameter growth and that the only variables in the equations are site index (S), age (A), total height (H), breast height diameter (DBH), and crown ratio (CR). In this set of variables we have variables that can describe tree size, vigor, competitive position and site productivity. We will not address the specifics of the appropriateness of variable selection, regression diagnostics or variable transformations except for the fact that this example uses a log transformation on some variables.

Our set of regression equations for diameter and height growth are,

$$\ln(DINC) = a_0 + a_1 \ln(DBH) + a_2 \ln(HINC) + a_3 CR + \epsilon_{DINC}$$
$$\ln(HINC) = b_0 + b_1 \ln(DBH) + b_2 H^2 + b_3 S + b_4 A + \epsilon_{HINC}$$

where ln is the natural logarithm and a_k and b_k are vectors of regression coefficients.

In this case, the model can be estimated using OLS, 2SLS, SUR or 3SLS. The model can be solved for the parameter estimates using various statistical software packages such as SAS (SAS Institute, Inc., 1987) and R (Ihaka and Gentleman, 1996). The R code, using the `systemfit` package (Hamann, 2002) to solve the equations looks like:

```
dinc <- ldg ~ ldbh + lhg + cr
hinc <- lhg ~ ldbh + ht2 + site + age
inst    <- ~ ldbh + cr + ht2 + site + age
labels <- list( "dinc", "hinc" )
system  <- list( dinc, hinc )

fit1sls <- ols.systemfit( system, inst, labels, data )
fit2sls <- twostage.systemfit( system, inst, labels, data )
fitsur <- sur.systemfit( system, inst, labels, data )
fit3sls <- threestage.systemfit( system, inst, labels, data )
```

For users of SAS `PROC MODEL`, to obtain parameter estimates, the code would look like:

```
proc model data=example;

        endogenous lhg;
        instruments ldbh cr ht2 site age;
```

```
        ldg = a0 + a1 * ldbh + a2 * lhg + a3 * cr;
        lhg = b0 + b1 * ldbh + b2 * ht2 + b3 * site + b4 * age;

        fit ldg lhg /ols;
        fit ldg lhg /sur;
        fit ldg lhg /2sls;
        fit ldg lhg /3sls;

run;
quit;
```

In this case, one would not use SUR to obtain parameter estimates because the equations contain endogenous variables. It is included here to demonstrate the syntax of the software package used to obtain estimates.

2.6 CONCLUSIONS

There are many choices for generating parameter estimates for systems of regression equations where the variables may or may not be related by the structural relationships or the error terms. Three-Stage Least Squares is appropriate when there are both endogenous variables and the equations are related through the error terms. Seemingly Unrelated Regressions can be used when the equations do not contain endogenous variables, but the error terms are related. The Two-Stage Least Squares method is used when the system of equations contain endogenous variables and the equations are not related through the error terms. If there are no endogenous variables and the equations are not related through the error terms, then Ordinary Least Squares is an appropriate method. For a detailed example of fitting a system of equations, including the determination of identification see Furnival and Wilson (1971). In short, the estimation method can be broken into a table, Table 2.1, that contains four cells, each cell represents a unique set of criteria that determines which estimation method may be used to obtain parameter estimates for a system of equations.

Type of Relationship	No Endogenous Variable	Endogenous Variables
Single Equation Estimation Methods	OLS	2SLS
Simultaneous Estimation Methods	SUR	3SLS

Table. 2.1: Decision matrix for determining which parameter estimation method is appropriate

The process of determining which method to otain parameter estimates is not always simple. If, over the course of model development, endogenous variables become insignificant and the system of equations changes classes the parameter estimates are equivalent to a method that utilizes less information. For example, if a system that is solved using 3SLS, has an endogenous variable that becomes insignificant and the model developer drops the term, or fits the models using SUR, the parameter estimates are equivalent. Likewise, if the model contains significant endogenous variables, but the variance-covariance matrix is a band-diagonal matrix (all of the off diagonal elements are zero) when fit using 3SLS, the system is equivalent to 2SLS. If there are no endogenous variables and there are no relationships in the error terms among the equations, then all the estimation methods are equivalent to OLS.

3. SIMULTANEOUS EQUATIONS FOR INDIVIDUAL TREE GROWTH IN YOUNG SOUTHERN OREGON AND NORTHERN CALIFORNIA DOUGLAS-FIR PLANTATIONS

3.1 ABSTRACT

This paper presents the results of a diameter and height increment model developed for Douglas-fir (*Psuedotsuga menzesii* (Mirb.) Franco) in young Southern Oregon and Northern California conifer plantations. The equations were fit using Ordinary Least Squares (OLS), Two-Stage Least Squares (2SLS) and Three-Stage Least Squares (3SLS) to examine the effects of fitting the equations independently and as a system of equations. The basal diameter growth model, which used height growth as an endogenous variable, was strongly influenced by the fitting method and was less robust when projected past the bounds of the data. The height growth model and the static crown ratio model were not strongly influenced by the other equations in the system. A few points at the extremes of the dataset appear to have had influence on the basal diameter growth model. Ordinary Least Square and Three-Stage Least Squares produced simlar projections when an example tree was projected sixty years into the future.

3.2 INTRODUCTION

The incorporation of young stand models into the forest planning process has benefits beyond simply projecting a tree list until conditions are appropriate for an inventory projection system to take over simulation of the older stand. Site preparation methods, stocking considerations, species composition, herbicide treatments, and stand density management decisions influence the development of young stands in such ways that make modeling the effects of those treatments within older stand models difficult. Typically, older stand models such as FVS (Stage, 1975), ORGANON (Hann et al., 1993), and CACTOS (Wensel et al., 1986) are designed to predict future stand conditions by projecting a tree list forward in time at intervals longer than those found in young stand models (Ritchie and Powers, 1993; Opalach et al., 1990). While these former models have proven to be adequate at projecting changes that occur over longer time intervals (5 to 10 years) in established stands, these models have limited capabilities in predicting the influence of management decisions that occur during stand establishment such as stand density manipulation or the removal of competing vegetation and are not designed to handle short term changes such as those found in the stand establishment time horizon which occurs from planting through fifteen years to twenty years.

Young plantation or regeneration models, used for those forest stands described as being less than 20 years old, have received less attention than older stand inventory models because young stands contain less value and are considered more structurally complicated than older stand models because of the influence of competing vegetation, difficulties in addressing site productivity and the influence of mortality (Ritchie and Powers, 1993; Hann et al., 1993; Wensel et al., 1986). Furthermore, these models are usually developed as a set of independent regression equations and are applied as such. These models often contain endogenous variables (Stage, 1975; Krumland and Wensel, 1981; Ritchie and Hann, 1985) such as height growth in diameter increment models and have been developed without addressing the "errors-in-variables" problems associated with fitting systems of equations that contain endogenous variables or variables that are predicted in one equation and used in another equation during the same projection period.

The development of growth equations for individual tree models typically includes the development of single equations to estimate growth rates for standard tree measurements. Historically, equations for diameter growth, height growth, crown recession, changes in stem taper, ingrowth and mortality have been developed independently and used together as a suite of equations that defined a growth model system (Hann et al., 1993; Ritchie and Powers, 1993; Stage, 1973b; Wensel et al., 1986). Typically, these equations contain a basic set of variables for the prediction of multiple attributes. Diameter at breast height, stem height, and some measure of the competitive position may be included as dependent variables in multiple equations to predict different attributes for the same observation such as diameter growth, height growth and crown recession. Unfortunately, developing cross-correlated equations with a common error structure may not capture the correct parameter estimates and may lead to prediction error because the error structure of biological data rarely contain an independent and normally distributed error term for any one of the individual equations. Contemporaneous correlation among equations and simultaneity bias among dependent variables can be addressed utilizing single equation methods of estimation such as Two-Stage Least Squares (2SLS) (Theil, 1953) and simultaneous estimation methods such as Three-Stage Least Squares (3SLS) (Zellner and Theil, 1962; Borders, 1989; Kmenta, 1997; Lemay, 1990). In addition, exogenous variables such as incoming photosynthetically active radiation (PAR), precipitation and topographical features such as slope, aspect and elevation can be treated accordingly within the system of equations without increasing the covariance among the dependent variables over independently developed equations.

Until recently, many of the models that address simultaneity bias and contemporaneous correlation have been whole stand models (Hasenauer et al., 1998; Gregoire, 1987). Hasenauer et al. (1998) used 2SLS and 3SLS in addition to Ordinary Least Squares (OLS) to examine the consequences of fitting a diameter increment, height increment, and crown ratio equation for individual trees in an older stand model for Norway spruce (*Picea Abies* L. Karst) in Austria. The authors stated the advantages of using 3SLS even when there are no endogenous variables with in a system as producing parameter estimates that were consistent and unbiased which would not be the case in models that were fit independently and used together as a system. They suggested, as do many econometrics texts (Kmenta, 1997; Greene, 2000), that the gain in effeciency will increase the precision of the model projections and the larger the cross equation correlations, the larger the increase in efficiency.

Hasenauer et al. (1998) and Rose and Lynch (2001) both addressed simultaneous equation methods in models that are applicable to older stands. Rose and Lynch (2001) developed Seemingly Unrelated Regression (SUR) models for basal area increment that utilized the correlation of the the stand basal area increment with the tree level basal area increment by utilizing a ranking method. The minimum age of the sample was 21 years. While age is rarely used in recently developed increment models, these authors have addressed the contemporaneous correlation problem by fitting models using a simultanous estimation method.

In young plantations, there may not be significant effects among increment equations because diameter growth, height increment and crown dynamics may not be influenced by each other as strongly as in older stands where inter-tree competition has a strong influence. The components of the equations may not be influenced by density or trees in a young stand may not have differentiated enough for a competitive position variable to be significant.

To date, there has been no work performed on examining the effects of simultaneous equation estimation for individual trees in young stand models. The paper examines the consequences of fitting individual tree increment models using least squares methods for individual equations and systems of equations for Douglas-fir in young plantations.

3.3 DATA

Data were collected from 109 plots established in stands in Northern California and Southern Oregon during the 1994 through 1999 field seasons. Sites were selected from plantations in National Forest, Bureau of Land Management and private land in Oregon and California. Candidate stands were selected from across a range of ages and elevations by stratifying all potential stands into age and elevation classes. All sites were sampled on a single two-year interval. The remeasurement was conducted two years, as near to the day as possible, subsequent to the initial measurement. In most instances, remeasurement was within two or three days of the scheduled two-year remeasurement date. The latitude of the selected stands ranged from 40° 10'59" N to 43°24'30" N. Longitude ranged from 123° 48'00" W to 121° 51'59" W.

All sites were plantations less than 26 years of age at the time of the initial measurement. Age was defined as the difference between the current year and the planting year. Some sites had residual trees from the previous stand. About 21 percent of the plots had greater than 20 square feet of basal area per acre on one or more plots in trees larger than 12 inches DBH. About 11 percent had greater than 20 square feet of basal area per acre on two or more plots.

A cluster of four to ten plots was established in each of the 109 candidate stands. Each plot consisted of a single fixed plot with an 11.78-foot (3.59 m.) radius for all trees with a DBH less than 6.1 inches (15.2 cm) and shrubs. Trees with a DBH larger than 6.1 inches were sampled using a 20 BAF variable-radius plot. For all woody vegetation with a basal diameter greater than 0.3 inches, D6 (basal diameter of the five largest stems 6 inches from the ground, measured to the nearest 0.1 inch using calipers); DBH (breast height diameter, measured to the nearest 0.1 inch, using calipers); HT (length of the longest stem, measured to the nearest 0.1 foot, using the tangent pole method); NSTEMS (total number of stems over 0.3 inches in diameter); HCB (height to base of the live crown, measured to the nearest 0.1 foot, using the tangent pole method); CWL (crown width measured on long axis, measured to the nearest 0.1 foot); CWS (short crown axis, measured to the nearest 0.1 foot); LEADER (length of the leader for the current year on trees shorter than 30 feet, measured to the nearest 0.1 foot); HT_INC (previous two-year internode length, measured to the nearest 0.1 foot); and a damage and severity code was measured and recorded. Two-year height growth (HINC) was calculated by subtracting the ending height from the initial height. Basal diameter increment (DINC) was computed as the change in D6 basal area translated back into a diameter increment.

Soil data were collected from a single soil pit in each stand located in the center of the plot cluster. Pit depth was determined by lithic contact. Field textural analysis was performed for each soil horizon. Water holding capacity (WHC) and percent rock content were calculated for each soil horizon and summed over all horizons to obtain total water holding capacity for the stand. Coefficients to calculate water-holding capacity for each layer were obtained from USDA (1984). Mean annual precipitation for stands in Southern Oregon were assigned from published soil association maps (Stearns-Smith and Hann, 1986). In Northern California mean annual precipitation was assigned using Isohyetal maps.

There were 1205 total observations in the original dataset. The final database, 428 observations, contained only those tree records that contained no damage, had positive height and basal diameter growth rates and were not dead in the remeasurement. The ranges for the variables are presented in Table 3.1.

Variable	Min	Mean	Max
Basal Diameter, (in)	0.10	2.040	17.90
Total Height, (ft)	2.01	9.139	53.40
Water Holding Capacity, (in)	2.78	6.594	11.25
Crown Ratio	0.18	0.810	0.98

Table. 3.1: Ranges for tree data.

3.4 METHODS

Since the objective of the study was to determine the advantages of using simultaneous methods of estimation for young tree increment models, both diameter and height models were developed as well as a static crown ratio model. Douglas-fir was the most abundant species in the dataset and was used to develop the models for investigation.

3.4.1 Dependent Variables

The change in basal diameter growth can be modelled in the same fashion as the chang in diameter at breast height. Modelling a change in diameter directly or computing a change in the diameter by transforming a change in basal area is widely used (West, 1980; Huang and Titus, 1995; Hann and Larsen, 1991; Dolph, 1992b; Stage, 1973b). West (1980) found no advantage of using basal area increment over diameter increment. The diameter increment model was developed from the basal diameter (D6) area increment (BINC) and then converted into a diameter increment (DINC) for the projections. Height increment (HINC) was estimated as the change in total height for undamaged trees. Both height and diameter increment were fit for the two year interval and then divided by two to obtain an annual projection.

3.4.2 Explanatory Variables

The explanatory variables, which typically fall into five categories Vanclay (1994), were included when applicable. The five classes are plant size, vigor, competitive position, density and site productivity. Since the data consisted mostly of plots that had not yet reached crown closure, some of the classes were not meaningful, primarily, density and competitive position.

The same classes of explanatory variables were used in the development of the basal diameter increment, height increment and crown ratio models. This was done to keep the model complexity to a minimum due to potential variable interaction.

3.4.2.1 Size

Tree size, commonly defined as diameter, height, volume or some combination thereof, was examined. In this case basal diameter (D6) and total height were examined in the model for analysis. Since DBH was only found on a subset of the plants, and only those trees with positive basal diameter and height growths were modelled, DBH was not included in the model for basal diameter growth.

3.4.2.2 Vigor

Crown ratio (Dolph, 1988a; Ritchie and Hann, 1990), diameter growth (Arney, 1985) and height growth (Wykoff et al., 1982) were examined as measures of tree vigor. Using predicted height growth as a vigor variable in the basal diameter function allowed the examination of an endogenous variable in the system of equations and thus when height growth was included in the diameter growth equation, the system was fit using 3SLS.

Height growth was also included in the basal diameter growth model to examine the effects of fitting the system using simultaneous estimation methods. Early model fits resulted in multicollinarity when basal diameter growth was included in the height growth model and height growth was included in the basal diameter growth model.

3.4.2.3 Competitive Position

The competitive position of a tree is often used to relate how well the plant will respond to release when given the opportunity (Vanclay, 1994). Basal area in larger trees is often used in inventory update models for older stands (Wykoff, 1990). Other authors have used basal area, broken into plant form classifications with some success (Ritchie and Powers, 1993). Basal area in taller (BAIN), as defined by the total sum of basal area in plants taller than the subject plant was examined as a potential competitive position variable. Crown ratio is often used in young models and since the variable is easy to compute when compared to basal area in larger trees it was chosen as a measure of tree vigor and competitive position (Dolph, 1992a). Basal area in larger is not a good measure in young stand as the majority of plants are not influenced by other plants though crown competition.

3.4.2.4 Density

Plot basal area (Dolph, 1992b; Hann and Larsen, 1991), stems per acre (Arney, 1985) and plot crown closure (Wykoff, 1990; Opalach et al., 1990) have successfully been used as measures of density. Most of the plots were far from crown closure and the influence of these variables were rarely significant. Plot percent crown closure was chosen to be the varible that would be included in the model because many authors believe a model should be robust enough to account for changes in density (Ritchie and Powers, 1993; Hann et al., 1993). None of these variables gave satisfactory results during the initial model development.

3.4.2.5 Site Productivity

Examined site productivty variables, those variables that allow a model to respond to different site productivity attributes, were water holding capacity, percent slope, aspect, and elevation. Water holding capacity was defined as the amount of available water found in the soil when saturated. These values are determined by measuring the texture of the soil horizons, determining the potential water holding capacity, and summing the potential water holding capacities for all the soil horizons. This variable is time intensive and more precise when compared to other tree-specific site productivity measures such as site index or site productivity index (Huang and Titus, 1995). Also, site index measurements in young stands are suspect as the estimates are subject to bias due to vegetation management patterns. Water holding capacity is advantagous because potential water holding capacity is independent of previous vegetation attributes as site index has been shown to be subject to shrub competition in early stages of stand establishment. Water holding capacity was also advantageous because there were no residual trees which could be utilized to estimate site index and site index has also been shown to be unreliable in young plantations (Newton and Hanson, 1998).

3.4.3 Equation Forms

The basic model form for the equations was intrinsically linear (Ratkowsky, 1990) so that the models could be linearized using logarithmic transformations. Since both height growth and basal diameter growth for Douglas-fir increase, peak, then decrease, a log-log model was used with a squared term and log term in the explanitory variables to mimic the behavoir of the expected response curve.

3.4.4 Parameter Estimation

The initial models were fit using linear regression in the R software package (Ihaka and Gentleman, 1996). An additional R package, presented in the Appendix (Hamann, 2002), was developed by the author to perform simultaneous equation estimation.

3.4.4.1 Initial Equation Development

The initial equations where developed by performing a stepwise regression on a basic set of variables. The same set of variables were included in the basal diameter and height increment functions so that variables that were significant to both equations would be included in both models when fit simultaneously. Only basal diameter (D6), total height (THT), water holding capacity (WHC), plot percent crown closure (PLOTPCC) and basal area in taller plants (BAT) where found to be significant for both of the initial models.

After the initial models were fit, the studentized residuals, standardized residuals were examined for both the basal diameter increment and height increment models. Influential observations were detected by examining plots of DFBETAS and DFFITS (Draper and Smith, 1998) for both models. There were intitially seven residual points for the height growth model that were outside the range of the rest of the residuals. The basal diameter growth model contained no outliers during the initial inspection.

The system of equations that was used for the final model analysis was,

$$\ln(BINC) = a_0 + a_1 \ln(D6) + a_2 \ln(HINC) + a_3 BAT \tag{3.1}$$

$$\ln(HINC) = b_0 + b_1 \ln(HT) + b_2 HT^2 + b_3 WHC + b_4 C \tag{3.2}$$

$$C = c_0 + c_1 \ln(D6) + c_2 HT^2 + c_3 BAT \tag{3.3}$$

where,

$$C = \ln\left(\frac{1}{CR} - 1\right) \tag{3.4}$$

The initial crown ratio model was developed as a logistic regression model which gave poor results. The model was then developed using a peaking function. It should be recognized that in young stands crown ratio is typically a monotonically decreasing function during stand establishment as trees are growing in open conditions and are subject to crown recession as crown closure occurs. These stands have not developed to a stage where crown recession would show as a response to release as would be the case in older stands where competition induced mortality or stand manipulation may occur.

3.4.4.2 Simultaneous Equation Development

These models were then fit using 2SLS to account for the endogenous variable, height growth, in the basal diameter growth and finally using 3SLS to account for the contemporaneous correlation. The models where fit using a custom computer program in R, `systemfit`, which is presented in the appendix. Diagnostics were then performed on the simultaneous fits and compared to the individual fits.

The residuals were plotted over the predicted values for basal area increment and height increment to ascertain the presence of heterscedasticity. Normal probability plots were also examined to determine normality. The predicted 2SLS and 3SLS values where plotted on the same graph to examine relationships between the residuals and the predicted between the two methods.

The cross-equation correlation, r_{ijk}, was computed for each of the predicted values from the three equations (Hasenauer et al., 1998). The cross-equation correlation was defined as,

$$r_{ijk} = \frac{x'_{ik} C_{ij} x_{jk}}{\sqrt{(x'_{ik} C_{ii} x_{ik})(x'_{jk} C_{jj} x_{jk})}} \tag{3.5}$$

where r_{ijk} is the correlation between the predicted values of equations i and j for the k^{th} observation. C_{ij} is the cross-equation variance-covariance matrix between equations i and j.

Improvement in prediction efficiency (Hasenauer et al., 1998),

$$\eta = \frac{se_{i,k}(2SLS)}{se_{i,k}(3SLS)} \tag{3.6}$$

where

$$se_{ik} = \sqrt{x'_{ik}\hat{C}_{ii}x_{ik}} \tag{3.7}$$

and x'_{ik} is the row vector of covariates for a single tree observation was computed for each of the observations in the database to ascertain the improvement of 3SLS over the 2SLS fitting method.

3.5 RESULTS

The root mean-squared-error (RMSE) and $\bar{R}^2$ values for the final models are presented in Table 3.2 and the parameter estimates and standard errors are presented in Table 3.3. A table of the equation correlations is presented in table 3.4.

Box plots for the residuals across the range of basal diameter and height classes for the diameter increment and height increment models were generated to examine the influence of fitting method on the results. The box plots for two year basal diameter increment and height increment residuals are presented in Figure 3.1.

The covariance matrices between the three equations were nonzero for all elements which shows correlation among the independent variables in the equations. The correlations among the predictions for the three equations are presented in Figure 3.2. No correlation among the predicted variables in

	OLS		2SLS		3SLS	
Model	RMSE	$\bar{R}^2$	RMSE	$\bar{R}^2$	RMSE	$\bar{R}^2$
BINC	0.72944	0.76293	0.72993	0.76261	0.72947	0.76291
HINC	0.56709	0.51537	0.56822	0.51343	0.56826	0.51336
LOGITCR2	0.61306	0.34843	0.61306	0.34843	0.61309	0.34838

Table. 3.2: Final model fit statistics for OLS, 2SLS and 3SLS methods

		OLS		2SLS		3SLS	
Model	Variable	Estimate	SE	Estimate	SE	Estimate	SE
BINC	(Intercept)	-5.360434	0.055979	-5.383557	0.086759	-5.352173	0.084726
	$\ln D6$	1.188661	0.053724	1.163077	0.090903	1.191516	0.088486
	$\ln HINC$	0.504312	0.055828	0.546578	0.133363	0.497897	0.128885
	BAT	-0.002965	0.000819	-0.002916	0.000831	-0.003102	0.000827
HINC	(Intercept)	-1.391300	0.138480	-1.397185	0.139425	-1.419886	0.138978
	$\ln HT$	0.815180	0.061414	0.774974	0.111570	0.776566	0.111267
	HT^2	-0.000495	0.000113	-0.000460	0.000139	-0.000463	0.000139
	WHC	0.048494	0.015575	0.046868	0.016053	0.049854	0.015954
	LOGITCR2	-0.194357	0.042803	-0.250117	0.136006	-0.250394	0.135547
LOGITCR2	(Intercept)	-1.518022	0.039921	-1.518022	0.039921	-1.519307	0.039864
	$\ln D6$	-0.674752	0.048128	-0.674752	0.048128	-0.669337	0.047385
	HT^2	0.000651	0.000112	0.000651	0.000112	0.000635	0.000108
	BAT	0.001924	0.000686	0.001924	0.000686	0.001987	0.000685

Table. 3.3: Final model estimates and standard errors for OLS, 2SLS and 3SLS methods

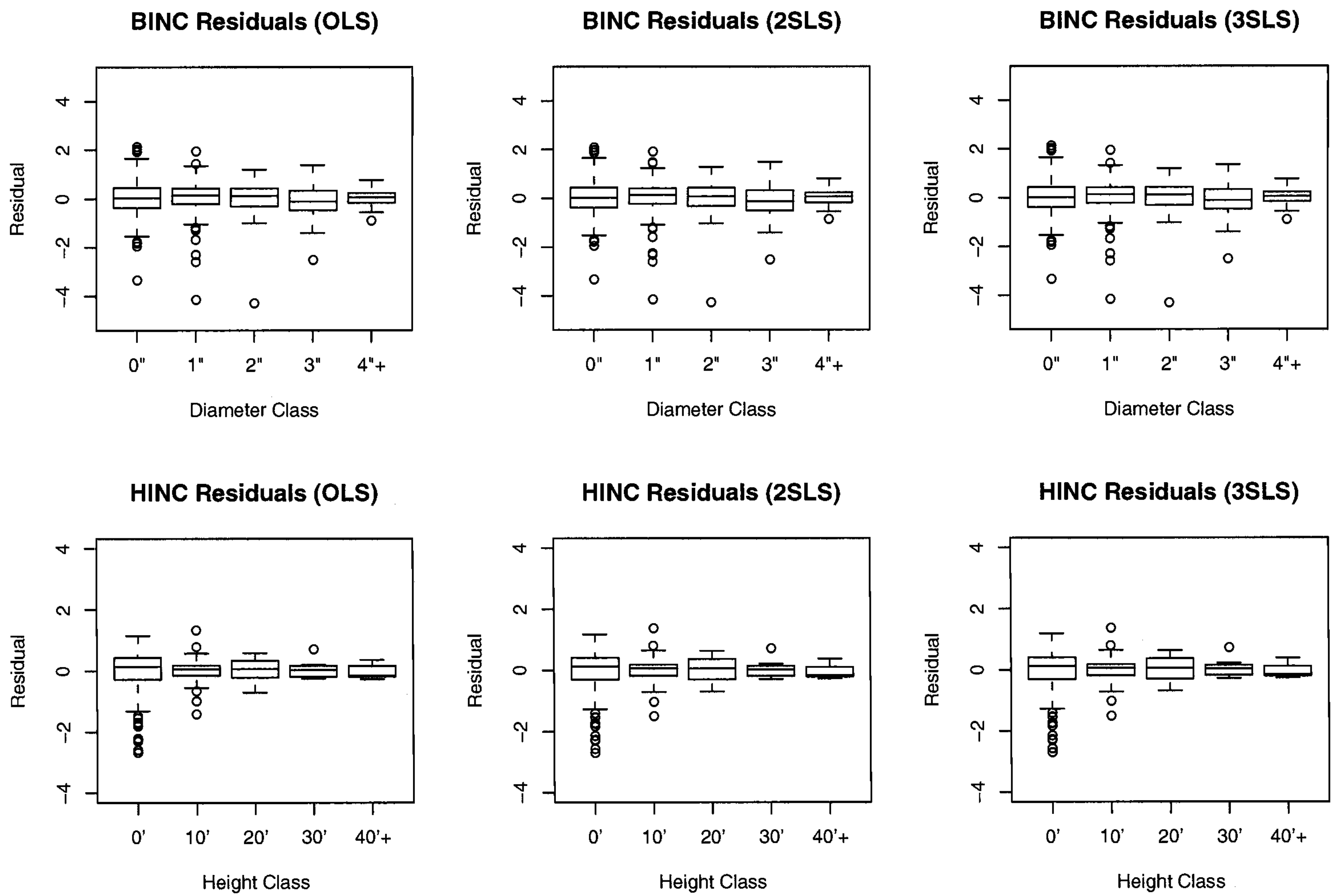

Figure. 3.1: Box plots for basal diameter growth and height growth over basal diameter and height

a set of equations should display a random pattern centered vertically about zero for the range of predicted values.

OLS			
	BINC	HINC	LOGITCR2
BINC	1.0000000	-0.06746815	-0.28518169
HINC	-0.0674681	1.00000000	0.00626702
LOGITCR2	-0.2851817	0.00626702	1.00000000
2SLS			
	BINC	HINC	LOGITCR2
BINC	1.000000	-0.1170135	-0.2773077
HINC	-0.117014	1.0000000	0.0651835
LOGITCR2	-0.277308	0.0651835	1.0000000
3SLS			
	BINC	HINC	LOGITCR2
BINC	1.0000000	-0.0809835	-0.2865397
HINC	-0.0809835	1.0000000	0.0666092
LOGITCR2	-0.2865397	0.0666092	1.0000000

Table. 3.4: Correlation matricies for OLS, 2SLS and 3SLS methods

3.6 DISCUSSION

The objectives of this study were to determine if using a simultaneous equation estimation method would produce superior parameter estimates and fit statistics over single equation methods such as 2SLS and to assess the behavior of a system of equations in long projections.

The answer to the first objective was simple. There was a slight increase in estimation efficiency by using a simultaneous approach to parameter estimation. The ratio in efficiency, as defined by equation (3.6), was computed

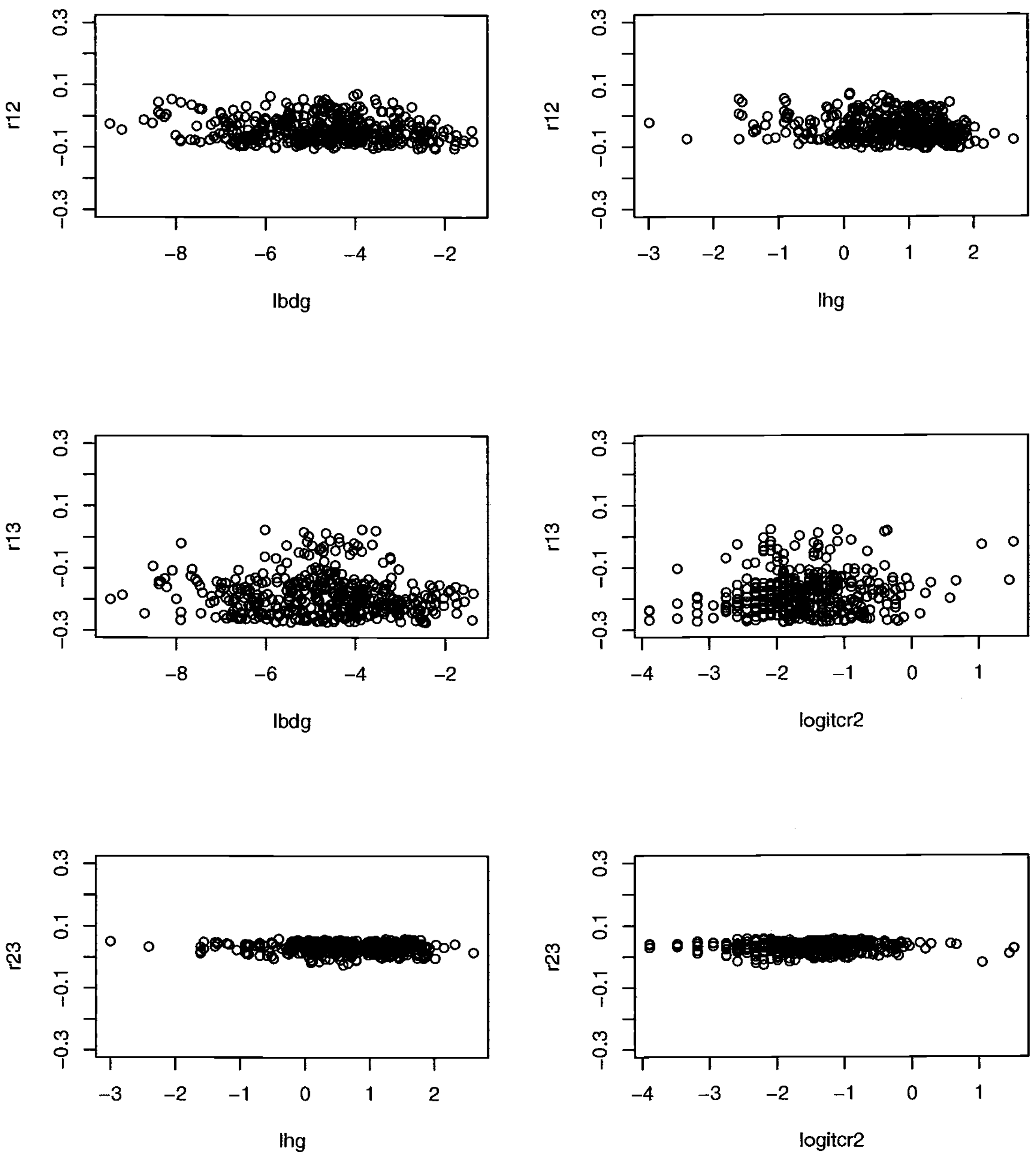

Figure. 3.2: The cross-equation correlations (r_{ij}) between each pair of predictions versus its predictions as the resulted from the 3SLS output. r12 indicates the correlation between BINC and HINC, r13 indicates the correlation between BINC and LOGITCR2 and r23 represents the correlation between HINC and LOGITCR2.

for all the observations in the database. Table 3.5 shows the range in increases in estimation efficiency for the 3SLS parameter estimates over the 2SLS.

	BINC	HINC	LOGITCR2
Min.	1.000	1.000	1.000
Mean	1.016	1.003	1.005
Max.	1.034	1.006	1.028

Table. 3.5: Prediction efficiency ratio (2SLS/3SLS) for final model (1.00 = no increase).

To examine the influence of simultaneous fitting methods on projected trees, a single tree was projected 60 years into the future without mortality. The initial tree values for basal diameter and total height were 0.1 and 1.0, respectively. These initial observations were considered typical of the young trees that would be projected using the set of equations. The models were projected using assuming 500 stems$\cdot$ac^{-1} and a water holding capacity of 7 which was slightly higher than the mean of the dataset. The tree record was assumed to be a dominant tree where the basal area in taller plants value was zero because a dominant tree would be less subject to mortality. Log bias was corrected for using the method described by Baskerville (1972).

3.6.1 Basal Diameter Increment

The $\bar{R}^2$ values for the basal area increment model remained roughly the same, as did the $\bar{R}^2$ values for the other models. The basal area increment function showed an improvement in RMSE when the fitting method accounted for contemporaneous correlation (3SLS) and show an improvement over the 2SLS when simultaneity bias was accounted for in the 3SLS fits. The RMSE values for the basal area increment function were 0.7294, 0.7300 and 0.7295 for the OLS, 2SLS and 3SLS methods, respectively. In this case, the RMSE value for the OLS fits suggests that model would produce the most precise estimates when compared to the 2SLS and 3SLS models. This may not be the case if there is a specification error by including the crown ratio variable in the height increment model. Hasenauer et al. (1998) found the crown ratio variable was insignificant in the 3SLS model and was thus removed from the system of equations, uncoupling the system of equations.

The significance of the parameter estimates in the basal diameter increment model was not strongly influenced by the fitting method as was the case in the height increment model. No terms became insignificant as the equations were fit simultaneously as other authors have found (Hasenauer et al., 1998). The standard error estimates for all terms but the basal area in taller term increased from about 0.05 to 0.08 and for the height growth term, increased from 0.05 to 0.12 when fit using a simultaneous method. The standard error estimates for the BAT term remained roughly the same value of 0.08 suggesting little simultaneity bias for the 2SLS and little evidence of cross-equation relationships for the 3SLS fits.

Basal diameter projections were carried out for 60 years, well past the time at which most young stand models are supposed to be useful to compare the effects of projecting young stands over long periods.

The OLS projected basal diameter at the end of the 60 year interval was 25.2 inches. The 2SLS projected basal diameter at the end of the 60 year projection was 23.4 or 92.8 percent of the OLS model. The projected basal diameter for the 3SLS equation was 25.60 or 102 percent of the OLS projections. The behavior of all three equations remained the same for the 60 year projection. All increased, peaked, then decreased at a decreasing rate. The OLS basal diameter growth model peaked at age 21 or 7.69 inches, the 2SLS model peaked at 20 years or 6.84 inches and the 3SLS model peaked at 21 years or 7.83 inches. All three models peaked near the same age, but the

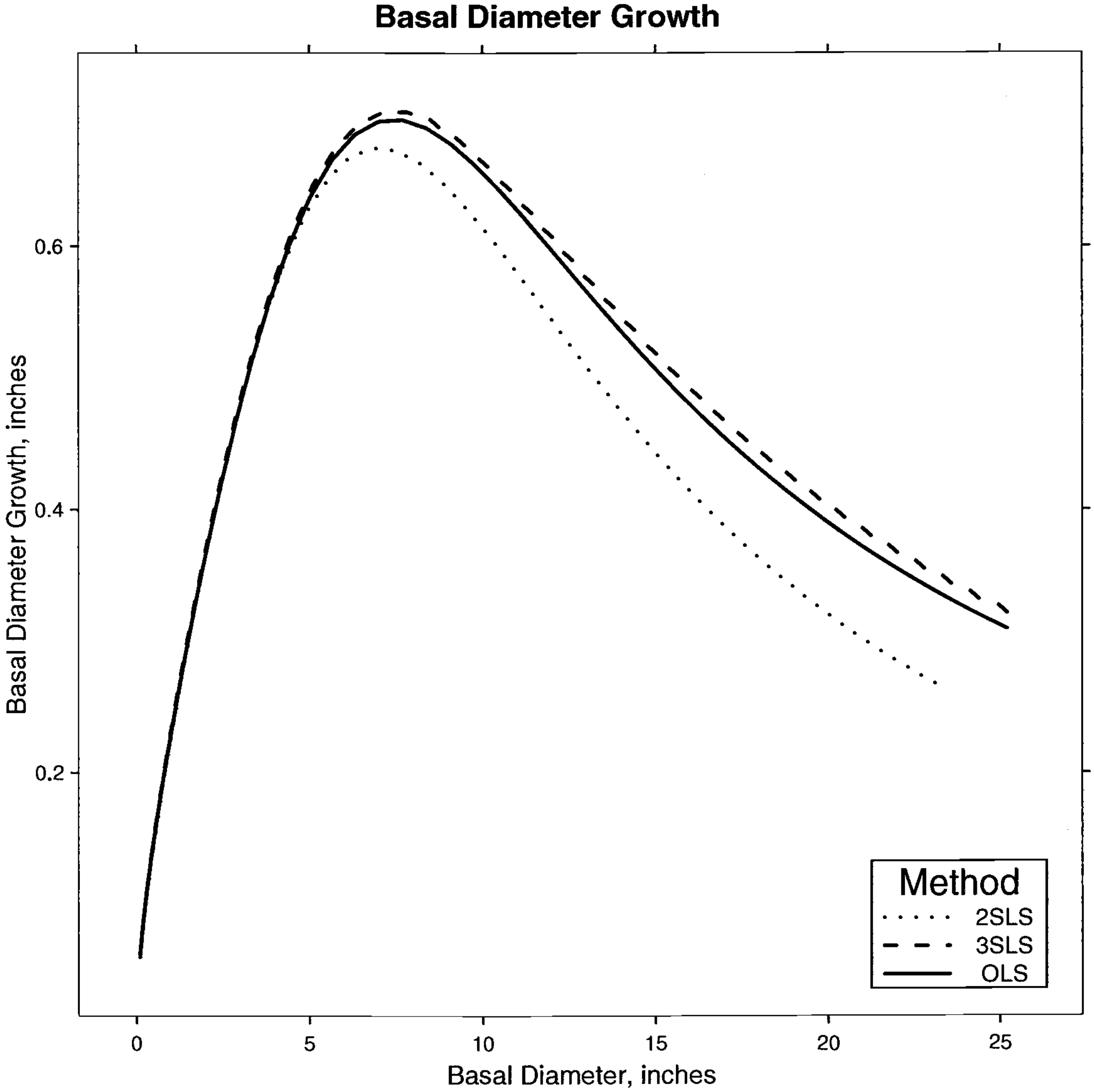

Figure. 3.3: Basal diameter increment over basal diameter for OLS/2SLS/3SLS models.

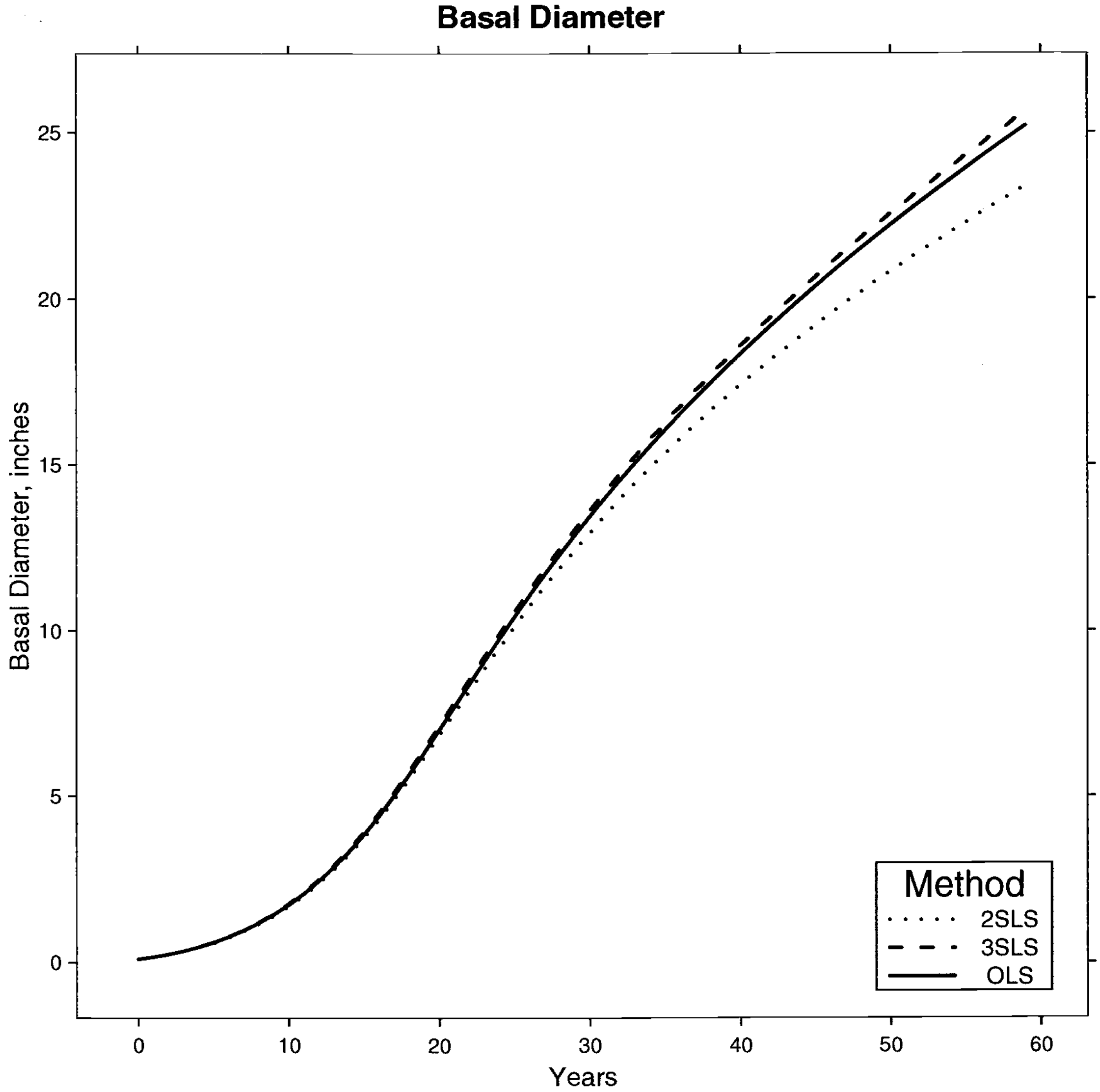

Figure. 3.4: Basal diameter increment projections for OLS/2SLS/3SLS models.

basal diameter values were very different in practical terms. The 3SLS and OLS models were roughly equivalent up to the peak at 20 or 21 years and had close to the same values in basal diameter and basal diameter growth. The 2SLS model peaked much sooner and the value was much less than either the OLS or 3SLS models.

Practically, the age at which young trees from small tree models are passed to older stand inventory projection models typically occurs before age 20 years. The projections showed no practical difference until the age of 30, after the peak in basal diameter increment. After 30 years, OLS produced higher basal diameters than did the 2SLS and 3SLS models and projections past age 30 should be used with caution regardless. In this case using either the OLS or 3SLS models should yield similar results until age 50, again well past the usefulness of a young tree growth model.

3.6.2 Height Increment

The changes in the height increment model were negligible as the values for the parameter estimates remained roughly the same for all three methods of estimation. The OLS and 2SLS parameter and standard error estimates for the height growth model remained the same and changed slightly when fit using the 3SLS method. All parameter estimates for the height growth model were significant.

The height increment model projection showed no preferable fitting method. Unlike the basal diameter increment model, there was no practical difference in the location of the peaks for all three fitting methods and very little difference in the projections resulting in similar projections for the entire 60 year period. All three fitting methods peaked at age 19 or 28 feet in height as can be seen in figure 3.5. The maximum height growth rates for the OLS,

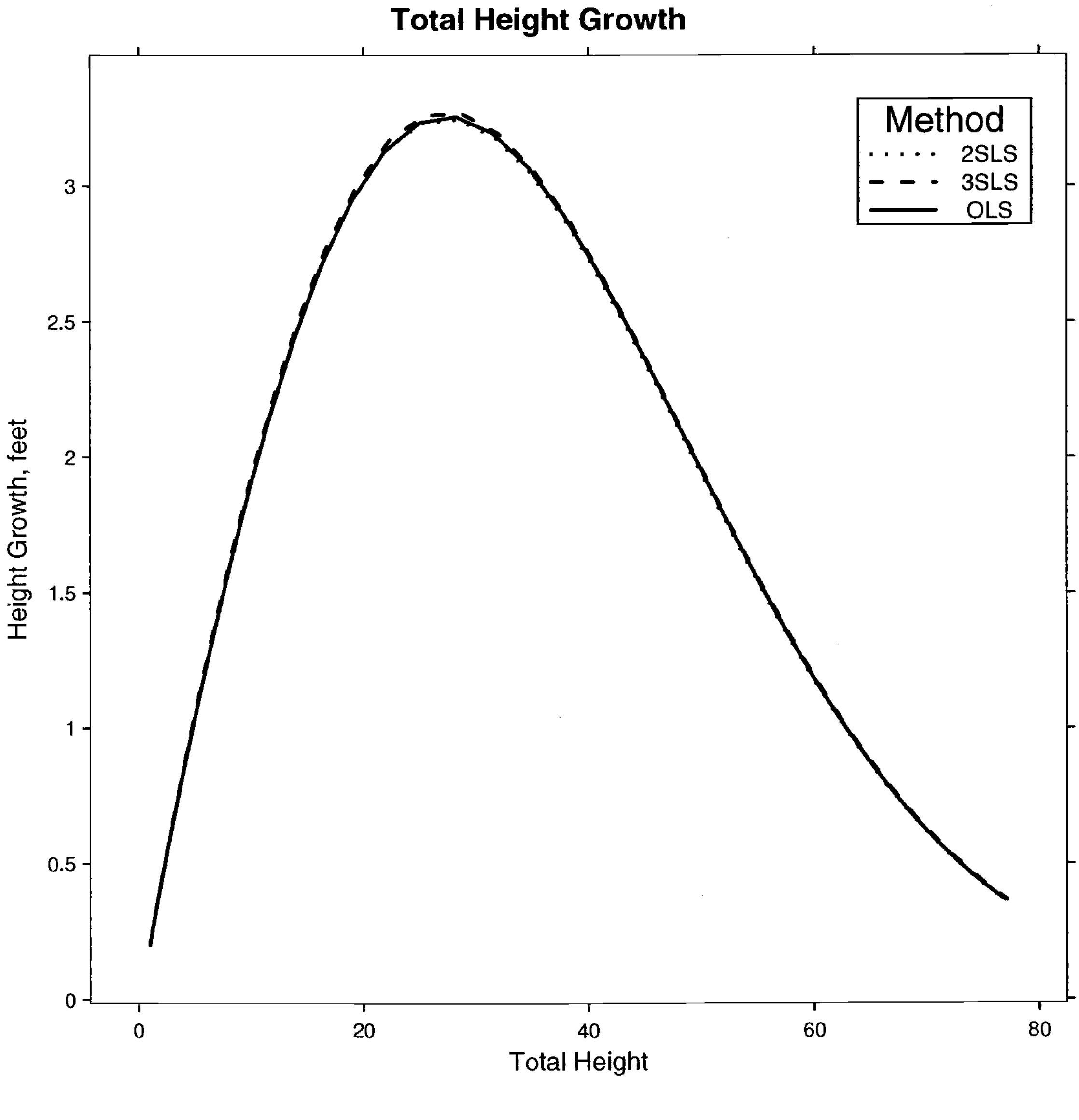

Figure. 3.5: Height growth over height for OLS/2SLS/3SLS models.

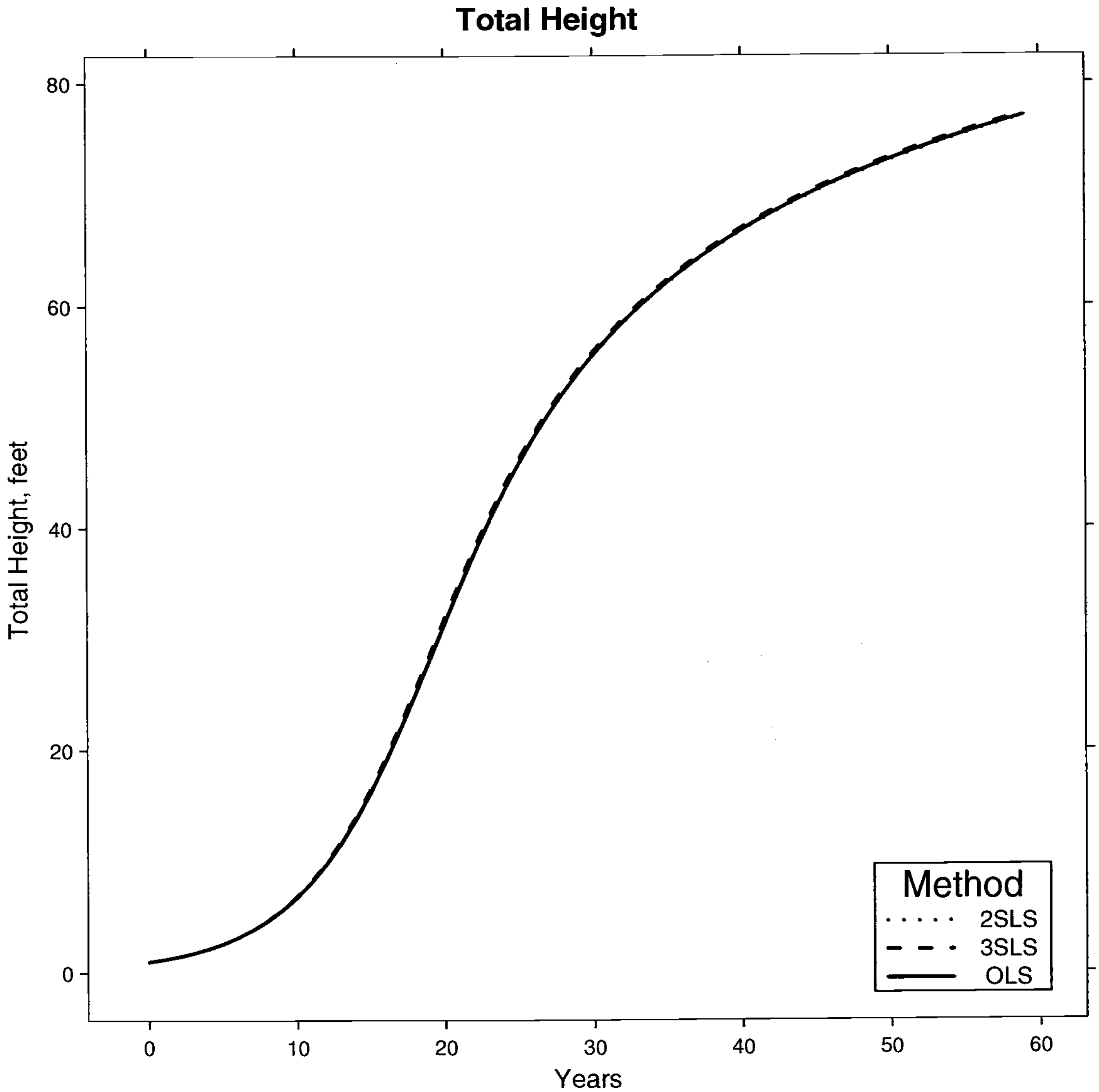

Figure. 3.6: Total height over time for OLS/2SLS/3SLS models.

2SLS and 3SLS methods were 3.26, 3.24 and 3.27 ft·yr^{-1}, respectively. The OLS model predicted a slightly lower total height at the end of the 60 year projection than the 2SLS and 3SLS models as can be seen in figure 3.6.

3.6.3 Crown Ratio

The crown ratio model had the same behavior as the height growth model for all three fitting methods. The $\bar{R}^2$ for all three fitting methods remained close to 0.348 for the crown ratio model.

As figure 3.7 shows, the fitting method had very little influence on the projections for the crown ratio model. The crown ratio for the OLS model was roughly the same until just after the peak at age 17 for all three fitting methods. The maximum crown ratio was about 0.91 for the OLS and 3SLS methods and 0.90 for the 2SLS method.

All three models would probably over-predict crown ratio when the projected trees are transfered to an older inventory projection type model since crown closure is not included in the models and mortality was not modelled here.

3.7 CONCLUSIONS

As discussed by other authors (Hasenauer et al., 1998), developing simultaneous regression equations is advantageous to single equation methods when growth models are based on multivariate attributes for a single tree observation. Simultaneous methods will produce consistent and efficient parameter and variance estimates over single equation methods even when there are no endogenous variables, if the equations are related in the disturbances.

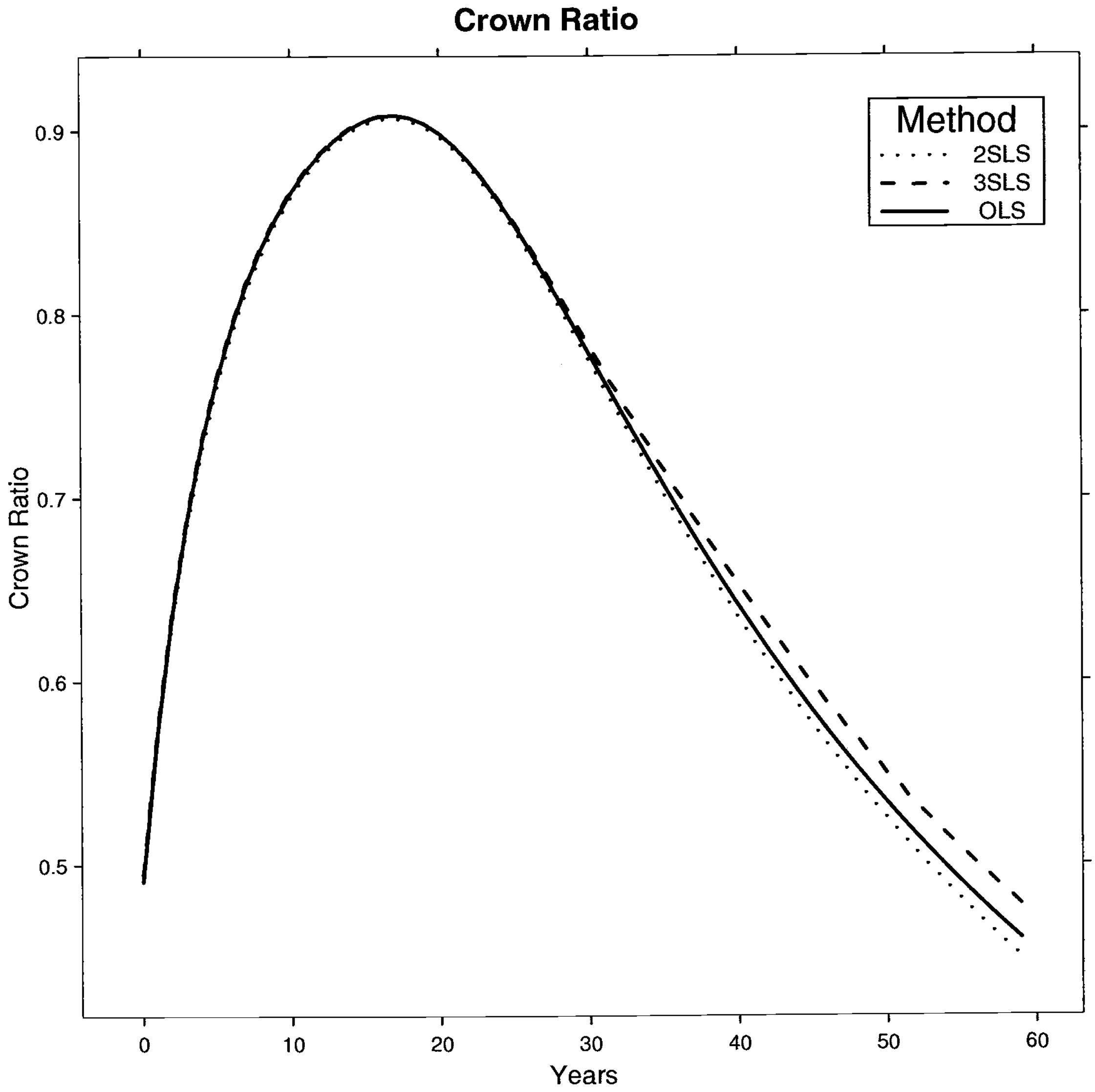

Figure. 3.7: Crown ratio over time for OLS/2SLS/3SLS models.

In this case, fitting the height growth and crown ratio models using a simultaneous technique yielded few if any advantages. The projections resulting from the OLS, 2SLS and 3SLS methods all yielded similar results except for the basal diameter increment model. The OLS and 3SLS models for basal diameter increment models behaved similarly. The values from projections for the 2SLS model were much less than anticipated and should not be used to project young trees to crown closure or some other age at which an older model would begin projecting the trees. The large difference between the 2SLS and the OLS and 3SLS projections is due to the differences in the D6 and HINC parameters in the three models. The D6 parameter is about 3 percent less in the 2SLS than the OLS or 3SLS models and the HINC parameter is about 5 percent higher in the 2SLS model than the OLS or 3SLS models. This result is most likely because the relationship between the error terms of the BINC and HINC equations is stronger than the endogenous relationship of the height growth parameter in the BINC model.

As most young stand models are not intended to be used past the range of data used to fit the models, a caution often disregarded, the models were projected out well past the time at which another model would project the tree records to examine the influence of inappropriately projecting tree records with a poorly fit model. In this case, if the user of the model would adhere to restricting the projections between ages one and fifteen or twenty, the OLS and 3SLS models would produce the similar results. In fact, all three models produce the same results until the basal diameter reaches about five inches which is typically the limit for young stand models.

Since the majority of the data for the model are below five inches in basal diameter, it is not surprising that all three fitting methods produce similar projections below five inch trees. The influence of the larger trees and the cross-equation correlation among the larger trees was significant by the results of the projections for all three models past the peak. The OLS model peaked later and had peaked at a larger value than the 3SLS model which would suggest that the OLS model may over-predict the maximum basal diameter growth as the parameter estimates for the OLS model is not asymptotically efficient or unbiased, unlike the 3SLS model.

While modellers often attempt to generate models that reflect the observations as accurately as possible, in many cases modelers will choose parameter estimates that are less efficient and biased over those that are not in order to obtain model projections that reflect their experience regardless of the model performance in terms of fit statistics or statistical test results.

The correlation of the residuals among the equations maintained the same signs for all three fitting methods. The correlation between the BINC and HINC and BINC and LOGITCR2 equations was negative, suggesting that there may be a change in allocation among the equations for the dataset as a function of tree size. The value of the correlation between the BINC and HINC changed from -0.07 for the OLS models to -0.11 for the 2SLS and back up to -0.08 for the 3SLS fits. The change was less noticable for the BINC and LOGITCR2 correlation.

It should be reinforced that most young stand models are not intended to be used past a certain size or age limit and since the models developed here had very few points over a basal diameter of five inches, the projections produced adequate results for the OLS and 3SLS models. If there were more observations in the large tree size classes the OLS and 3SLS models would probably be more similar at and past the peak in the projections.

The parameter estimates, RMSE values and standard error estimates for the final 3SLS are unbiased and consistent barring a specification error from one of the equations. Both the single equation and system methods showed no bias over the range of basal diameter and heights in terms of the box plots patterns.

The projections in the figures are based on a dominant tree (BAT = 0.0).

While young stands might not be subject to crown competition, the BAT term, does influence the height growth and thus the basal diameter increment predictions as does the site productivity variable, water holding capacity. Since the coefficients and magnitude for the BAT and WHC parameters are similar for all three fitting methods, the values for the height growth and subsequent basal diameter increment values scale accordingly.

As the population of young stand models grows, it is advisable to consider developing systems of equations using a simultaneous fitting method to account for contemporaneous correlation and if applicable, simultaneity bias among equations. As was the case in this study, the differences in projections are small for the equations that are not subject to contemporaneous correlation especially for ages below older stand models are considered better suited for projecting stands. When equations are subject to simultaneity bias, as was the case in the basal diameter increment model, using a simultaneous fitting method provides an increase in prediction efficiency that may be justified.

4. CONCLUSIONS

The fitting methods analyzed during this study provide advantageous techniques for fitting of systems of equations. These methods incorporate additional information that other methods fail to include such as the relationships among the variables in different equations and the correlation among the error terms. The list examined in this document is by no means exhaustive. The method that gave the best results did however include the most information among the three fitting methods examined.

4.1 MODEL VARIABLES

The variables used to develop the models in these manuscripts were of a simple nature and are commonly used in forest growth models. These variables are by no means exhaustive and as statistical models further develop and are combined with theoretical models, the variables may become more complex or surprisingly simple as the ability to confidently collect these variables increases.

In addition to the limited variables used in the model development, this system of equations only included three equations. Many models that predict individual tree attributes require many times more variables. Crown ratio, mortality, DBH, stem taper, crown profile and foliage characteristics are commonly found in established stand models. Models that address the issues associated with established stands may provide a more fertile environment for studying the influence of simultaneous equation estimation methods as these stands exhibit more complex growth patterns both horizontally and vertically throughout the stand because of differences in species, density and management.

Of course more data would have been nice. The majority of the data in this study occurred below five inches basal diameter which is the lower limit of crown closure which often occurs before the young stands have had time to undergo crown closure. The dynamics found in established stand models are more complex. It would be interesting to apply these methods to current models that address established stands which were developed using independent regression equations, and observe the differences between the two models.

4.2 FITTING METHODS

There are many other fitting methods available which may or may not produce better results. These three methods are the most common and were chosen for that reason alone. As the 2SLS and 3SLS methods are becoming more popular in statistical packages, more models will be fit using these methods. As the results may be more statistically sound in terms of efficiency and consistency, they may not provide the best solution to the problem of modeling the development of a young forest. In addition to producing meaningful results, models for young forests must meet an additional criteria older stand models are not subject to which is the ability to closely match the results of other models at the upper end of the projection.

In this case, OLS and 3SLS provided similar results for the equations that did not contain endogenous variables. The parameter estimates for the basal diameter increment function, which contained a simple relationship with the height growth function, were influenced by the estimation method.

A potential drawback to using a simultaneous method should be mentioned here. Most common software packages do not allow unbalanced data. That is to say that a dataset cannot contain missing values for endogenous variables. In forestry data sets for example, heights are rarely measured on every tree and for a model that contains both height and diameter growth equations, only a subset of the diameter growth obervations will be used because the corresponding height growth is not present. The effect is to lower the effective number of observations and thus, reducing the degrees of freedom in the model.

4.3 PROJECTIONS

The projections in this study were relatively simplistic in that only one tree record was used and mortality was not introduced over the sixty year projection. In short, the tree used for the results and discussion was a dominant tree without competition. The conclusions regarding the fitting methods were the same that other authors had found regardless (Hasenauer et al., 1998).

Had mortality been included in this study, the results for basal diameter projections may have been different. Again, this speculation is for naught as the database contained one two-year remeasurement and the number of observations in the final analysis was too low for a mortality model to be developed.

As with other studies, the increase in prediction efficiency was below 10 percent and almost undetectable for the crown ratio model. While these results may seem small, the influence over many projections would prove to be highly advantageous. Since future values are the sum of the previous predictions plus the predicted growth, using a system approach to estimation with lagged endogenous variables may show impressive gains.

4.4 R PACKAGE DEVELOPMENT

The R software package is a constantly evolving open source software project for statistical computing and graphics. Unlike SAS (SAS Institute, Inc., 1987), R is available without cost and the source code is freely available for end users to examine and modify at will.

Based on the S system, R is both a language and an application which allows users to develop additional modules to meet specific needs not built directly into the original package. These modules, or packages as they are referred to in R, can be downloaded from the Internet and installed locally. These packages are developed by anyone interested in extending R and can be submitted to the project for inclusion in further releases. The `systemfit` package (Hamann, 2002) was developed to perform simultaneous equation estimation within the R environment as well as specific data analysis used in this manuscript.

The `systemfit` package was developed in the R language as a package that is capable of fitting three types of systems of equations. The three methods of estimation are OLS, 2SLS, and 3SLS. The package is available from the author or directly for download from the R project site (www.r-project.org). The package was succuessfully verified against published examples (Kmenta, 1997) and against SAS. Further developments for `systemfit` should include the ability to address SUR, non-linear models and unbalanced datasets.

BIBLIOGRAPHY

Amemiya, T. (1977). The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica*, 45(4):955–968.

Arney, J. (1973). Tables for quantifying competitive stress on individual trees. Informal Report BC-X-78, Canadian Forest Service, Pacific Research Center, Victoria, British Columbia.

Arney, J. D. (1985). A modeling strategy for the growth projection of managed stands. *Canadian Journal of Forest Research*, 15:511–518.

Baskerville, G. L. (1972). Use of logarithic regression in the estimation of plant biomass. *Canadian Journal of Forest Research*, 2:49–53.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, New York, New York.

Bella, I. E. (1971). A New Competition Model of Individual Trees. *Forest Science*, 17(3):364–372.

Biging, G. S. and Wensel, L. C. (1985). Site index equations for young growth mixed conifers of Northern California. Research Note 8, Northern California Forest Yield Cooperative, Department of Forestry and Resource Management, University of California, Berkeley, California.

Borders, B. E. (1989). Systems of equations in forest stand models. *Forest Science*, 35(2):548–556.

Borders, B. E. and Bailey, R. L. (1986). A compatible system of growth and yield equations for slash pine fitted with restricted three-stag least squares. *Forest Science*, 32(1):185–201.

Botkin, D. B., Janak, J. F., and Wallis, J. R. (1972). Some ecological consequences of a computer model of forest growth. *Journal of Ecology*, 60:49–872.

Box, M. J. (1971). Bias in Nonlinear Estimation. *Journal of the Royal Statistical Society*, 33:171–202.

Brand, D. G. (1986). A competition index for predicting the vigor of planted Douglas-fir in southwestern British Columbia. *Canadian Journal of Forest Research*, 16(1):23–29.

Burkhart, H. E., Farrar, K. D., Amateis, R. L., and Daniels, R. F. (1987). Simulation of Individual Tree Growth and Stand Development in Loblolly Pine Plantations on Cutover, Site Prepared Areas. Bulletin FWS-1-87, School of Forestry and Wildlife, Virgina Polytechnic Institute and State University, School of Forestry and Wildlife, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

Burton, P. J. (1993). Some limitations inherent to static indices of plant competition. *Canadian Journal of Forest Research*, 23:2141–2152.

Cannell, M. G. R., Rothery, P., and Ford, E. D. (1984). Competition within stands of picea sitchensis and pinus contorta. *Annals of Botany*, 53:349–362.

Chambers, C. J. (1980). Empirical Growth and Yield Tables for the Douglas-Fir Zone. Report 41, Department of Natural Resources, Olympia, Washington.

Curtis, R. O. (1967). Height-Diameter and Height-Diameter-Age Equations for Second-Growth Douglas-Fir. *Forest Science*, 13(4):365–375.

Curtis, R. O. (1982). A Simple Index for Stand Density. *Forest Science*, 28(1):92–94.

Curtis, R. O., Clendenen, G. W., and DeMars, D. J. (1982). A new stand simulator for coast Douglas-fir: DFSIM user's guide. General Technical Report PNW-135, Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture, Portland, Oregon.

D'Agostino, R. B. (1971). An omnibus test of normality for moderately large size samples. *Biometrika*, 58:341–348.

D'Agostino, R. B. and Pearson, E. S. (1973). Tests for departures from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*, 60:613–622.

D'Agostino, R. B. and Tietjen, G. L. (1973). Approaches to the null distribution of $\sqrt{b_1}$. *Biometrika*, 60:169–173.

Daniels, R. F. and Burkhart, H. E. (1988). An integrated system of forest stand models. *Forest Ecology and Management*, 999(23):159–177.

Dolph, L. K. (1987). Site index curves for young-growth California white fir on the western slopes of the Sierra Nevada. *Research Paper PSW-185. Berkeley, CA: Pacific Southwest Forest and Range Experiment Station, Forest Service, US Dept. Agriculture.*

Dolph, L. K. (1988a). Predicting height increment of young-growth mixed conifers in the Sierra Nevada. Research Paper PSW-RP-191, USDA Forest Service, Pacific Soutwest Research Station, Berkeley, California.

Dolph, L. K. (1988b). Prediction of Periodic Basal Area Increment for Young-Growth Mixed Conifers in the Serria Nevada. Research Paper PSW-190, United States Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California.

Dolph, L. K. (1991). Polymorphic site index curves for red fir in California and southern Oregon. Research Paper PSW-RP-206, USDA Forest Service, Pacific Soutwest Research Station, Berkeley, California.

Dolph, L. K. (1992a). A diameter increment model for red fir in california and southern oregon. Research Paper PSW-RP-210, USDA Forest Service, Pacific Soutwest Research Station, Berkeley, California.

Dolph, L. K. (1992b). Predicting height increment of young-growth red fir in california and southern oregon. Research Paper PSW-RP-214, USDA Forest Service, Pacific Soutwest Research Station, Berkeley, California.

Donnelly, D. M. (1997). Pacific Northwest Coast Variant of the Forest Vegetation Simulator. Technical report, WO-Forest Management Service Center, USDA Forest Service, Fort Collins, Colorado.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. Jonh Wiley and Sons, New York, 3rd edition edition.

Drew, T. J. and Flewelling, J. W. (1979). Stand Density Management: An Alternative Approach and Its Application to Douglas-fir Plantations. *Forest Science*, 25(3):518–532.

Durbin, J. (1970). An alternative to the bounds test for testing serial correlation in least squares regression. *Econometrika*, 38:422–429.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. *Biometrika*, 37:409–428.

Ek, A. R. and Monserud, R. A. (1974). FOREST: A computer model for simulating the growth and reproduction of mixed species forest stands. *University of Wisconsin, School of Natural Resources. Research Report*, A2635:85 p.

Enquist, B. J., Brown, J. H., and West, G. B. (1998). Allometric scaling of plant energetics and population density. *Nature*, 395:163–165.

Farr, W. A., DeMars, D. J., and Dealy, J. E. (1989). Height and crown width related to diameter for open-grown western hemlock and Sitka spruce. *Canadian Journal of Forest Research*, 19:1203–1207.

Ferguson, D. E. and Adams, D. L. (1980). Response of advanced grand fir regeneration to overstory removal in Northern Idaho. *Forest Science*, 26(4):537–545.

Flewelling, J. W. and Jong, R. D. (1994). Considerations in simultaneous curve fitting for repeated height-diameter measurements. *Canadian Journal of Forest Research*, 24:1408–1414.

Ford, E. D. and Diggle, P. J. (1981). Competition for light in a plant monoculture modelled as a spatial stochastic process. *Annals of Botany*, 48:481–500.

Furnival, G. M. and Wilson, R. W. (1971). Systems of Equations for Predicting Forest Growth and Yield. In Patil, G. P., Pielou, E. C., and Waters, W. E., editors, *Statistical Ecology*, volume 3, pages 43–55. Penn State University Press, University Park.

Gallant, A. R. (1975). Seemingly Unrelated Nonlinear Regression. *Journal of Econometrics*, 999(3):35–50.

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation. *Biometrika*, 27:310–332.

Gill, S. J., Biging, G. S., and Murphy, E. C. (2000). Modeling confier tree crown radius and estimating canopy cover. *Forest Ecology and Management*, 126:405–416.

Goldfeld, S. M. and Quandt, R. E. (1965). Some tests for homoskedasticity. *Journal of the American Statistical Association*, 60:539–547.

Gorham, E. (1979). Shoot height, weight and standing crop in relation to density of monospecific plant stands. *Nature*, 279:148–150.

Greene, W. H. (2000). *Econometric Analysis:Fourth Edition.* Prentice Hall, Upper Saddle River, New Jersey, 4th edition edition.

Gregoire, T. G. (1987). Generalized error structure for foresty yield models. *Forest Science*, 33:423–444.

Hamann, J. D. (2002). `systemfit`:An R Package. Hamann, Donald and Associates, Inc., PO Box 1421, Corvallis, Oregon 97339.

Hann, D. W., , and Scrivani, J. A. (1987). *Dominant-Height-Growth and Site-Index equations for Douglas-fir and Ponderosa pine in Southwest Oregon*, volume 59. Forest Research Laboratory, Oregon State University, Corvallis, Oregon.

Hann, D. W. and Larsen, D. R. (1991). Diameter growth equations for fourteen tree species in southwest oregon. *Forest Research Laboratory, Oregon State University, Corvallis. Research Bulletin*, 69.

Hann, D. W., Olsen, C. L., and Hester, A. S. (1993). *ORGANON: User's Manual.* Department of Forest Resources, College of Forestry, Oregon State University, Corvallis, Oregon, 1.2 edition.

Hann, D. W. and Ritchie, M. W. (1988). Height growth rate of douglas-fir: A comparison of model forms. *Forest Science*, 34(1):165–175.

Hasenauer, H., Monserud, R. A., and Gregoire, T. G. (1998). Using Simultaneous Regression Techniques with Individual-Tree Growth Models. *Forest Science*, 44(1):87–95.

Hawksworth, F. G. (1977). The 6-class dwarf mistletoe rating system. Gen. Tech. Rep. RM-48, Rocky Mountain Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture, Fort Collins, CO.

Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *Journal of the Royal Statistical Society*, 47:103–114.

Huang, S. and Titus, S. J. (1993). An index of site productivity for uneven aged and mixed-species stands. *Canadian Journal of Forest Research*, 23:558–562.

Huang, S. and Titus, S. J. (1994). An age-independent tree height prediction model for boreal spruce-aspen stands in alberta. *Canadian Journal of Forest Research*, 24:1295–1301.

Huang, S. and Titus, S. J. (1995). An individual tree diameter increment model for white spruce in Alberta. *Canadian Journal of Forest Research*, 25:1455–1465.

Huang, S. and Titus, S. J. (1999). Estimating a system of nonlinear simultaneous individual tree models for white spruce in boreal mixed-species stands. *Canadian Journal of Forest Research*, 29:1805–1811.

Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.

Intriligator, M. D. (1978). *Econometric Models, Techniques, and Applications.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632.

Kelejian, H. H. (1971). Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables. *Journal of the American Statistical Association*, 66(334):373–374.

King, J. E. (1966). Site index curves for Douglas-fir in the Pacific Northwest. *Weyerhaeuser Co. Centralia, Washington. Weyerhaeuser Forestry Paper*, 8.

Kmenta, J. (1997). *Elements of Econometrics.* University of Michigan Publishing Publishing, Ann Arbor, Michigan, 2nd edition.

Krajicek, J. E., Brinkman, K. A., and Gengrich, S. F. (1961). Crown Competition - A Measure of Density. *Forest Science*, 7(1):35–42.

Krumland, B. and Wensel, L. C. (1981). A Tree Increment Model System for Northern Coastal California. Research Note 15, Co-op Redwood Yield Research Project, Department of Forestry and Conservation, College of Natural Resources, University of California, Berkeley, California 94720.

Larsen, D. R. and Hann, D. W. (1987). Height-Diameter Equations for Seventeen Tree Species in Southwest Oregon. Research Paper 49, Forest Research Laboratory, College of Forestry, Oregon State University, Corvallis, Oregon 97330.

Lemay, V. M. (1990). MSLS: a linear least squares technique for fitting a simultaneous system of equations with a generalized error structure. *Canadian Journal of Forest Research*, 20:1830–1839.

Levene, H. (1960). Robust tests for equality of variances. In Olkin, editor, *Contributions to Probability and Statistics*, pages 278–292. Stanford University Press.

Martin, G. L. and Ek, A. R. (1984). A Comparison of Competition Measures and Growth Models for Predicting Plantation Red Pine Diameter and Height Models. *Forest Science*, 30(3):731–743.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). *Applied Linear Statistical Models*. McGraw-Hill, Chicago, 4 edition.

Newton, M. and Hanson, T. J. (1998). Bias in site estimation from early competition. In *19th Forest Vegetation Management Conference*, pages 78–84, Redding, California. Forest Vegetation Management Conference, Forest Vegetation Management Conference.

Nystrom, K. and Kexi, M. (1997). Individual tree basal area growth models for young stands of Norway spruce in Sweden. *Forest Ecology and Management*, 999(97):173–185.

Opalach, D., Wagner, R. G., Maxwell, B. D., Dukes Jr., J. H., and Radosevich, S. R. (1990). A Growth model for young Douglas-fir stands: Progress report No. 2. *Forest Research Laboratory, Oregon State University, Corvallis, Oregon.*

Paine, D. P. and Hann, D. W. (1982). Maximum Crown-Width Equations for Southwestern Oregon Tree Species. Research Paper 46, Forest Research Lab, School of Forestry, Oregon State University, Corvallis, Oregon 97331.

Parke, W. R. (1982). An Algorithm for FIML and 3SLS Estimation of Large Nonlinear Models. *Econometrica*, 50(1):81–95.

Parks, R. W. (1967). Efficient Estimation of a System of Seemingly Unrelated Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated. *Journal of the American Statistical Association*, 62(318).

Pienaar, L. V. and Turnbull, K. J. (1973). The Chapman-Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. *Forest Science*, 19:2–22.

Powers, R. F. and Oliver, W. W. (1978). Site classification of ponderosa pine stands under stocking control in California. Research Paper PSW-128, Pacific Southwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture. 9 p.

Prentice, R. L. (1974). Degrees-of-freedom modifications for F tests based on non-normal errors. *Biometrika*, 61:559–563.

Puettmann, K. J., Hann, D. W., and Hibbs, D. E. (1993). Evaluation of the Size-Density Relationships for Pure Red Alder and Douglas-Fir Stands. *Forest Science*, 39(1):7–27.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models*, volume 107 of *Statistics, textbooks and monographs.* Marcel Dekker, New York.

Reineke, L. H. (1933). Perfecting a Stand-Density Index for Even-Aged Forests. *Journal of Agricultural Research*, 46(7):627–638.

Ritchie, M. W. and Hann, D. W. (1985). Equations for predicting basal area increment in Douglas-fir and grand fir. Research Bulletin 51, Forest Research Laboratory, Oregon State University, Corvallis.

Ritchie, M. W. and Hann, D. W. (1986). Development of a tree height growth model for Douglas-fir. *Forest Ecology and Management*, 15:135–145.

Ritchie, M. W. and Hann, D. W. (1990). Equations for predicting the 5-year height growth of six conifers species in southwestern Oregon. *Forest Research Laboratory, Oregon State University, Corvallis*, Research Bulletin 54:14.

Ritchie, M. W. and Powers, R. F. (1993). Users guide for SYSTUM-1 (version 2.0): A simulator of growth trends in young stand under management in California and Oregon. General Technical Report PSW-GTR-147, Pacific Southwest Forest and Range Experiment Station, Forest Service, US Dept. Agriculture, Berkeley, CA.

Rose, C. E. and Lynch, T. B. (2001). Estimating parameters for tree basal area growth with a system of equations and seemingly unrelated regressions. *Forest Ecology and Management*, 148:51–61.

SAS Institute, Inc. (1987). *SAS User's Guide.* SAS Institute, Inc., Cary, NC.

Schroder, J. and von Gadow, K. (1999). Testing a new competition index for Maritime pine in northwestern Spain. *Canadian Journal of Forest Research*, 29:280–283.

Schumacher, F. X. and Coile, T. S. (1960). *Growth and Yields of Natural Stands of Southern Pines.* T. S. Coile, Inc., Durham, N.C.

Scott, W., Meade, R., Leon, R., Hyink, D., and Miller, R. (1998). Planting denisty and tree-size relations in coast Douglas-fir. *Canadian Journal of Forest Research*, 28:74–78.

Shapiro, L. R., Wilk, M. B., and Chen, H. J. (1968). A comparate study of various tests for normality. *Journal of the American Statistical Society*, 63:1343–1372.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611.

Smith, S. H. and Bell, J. F. (1983). Using competitive stress index to estimate diameter growth for thinned douglas-fir stands. *Forest Science*, 29(3):491–499.

Stage, A. R. (1973a). A mathematical approach to polymorphic site index curves for grand fir. *Forest Science*, 9(2):167–180.

Stage, A. R. (1973b). Prognosis model for Stand Development. General Technical Report INT-137, USDA Forest Service Research Note, Ogden, Utah.

Stage, A. R. (1975). Prediction of height increment for models of forest growth. Research Note INT-I64, Intermountain Forest and Range Experiment Station, Forest Service, US Department of Agriculture, Ogden, UT.

Stage, A. R. (1976). An Expression for the Effect of Aspect, Slope, and Habitat Type on Tree Growth. *Forest Science*, 22(4):457–760.

Stearns-Smith, S. C. and Hann, D. W. (1986). Forest soil associations of southwest oregon. *Forest Research Laboratory, Oregon State University, Corvallis.*

Szwagrzyk, J. (1997). Modelling competition among trees in mixed stands of complex structure. *Zeszyty Naukowe. Akademii Rolniczej im. H. Kollataja w Krakowie. Rozprawy nr.*

Tang, S., Yong, L., and Wang, Y. (2001). Simultaneous equations, error-in-variable models, and model integration in systems ecology. *Ecological Modelling*, 142:285–294.

Theil, H. (1953). Repeated Least Squares applied to complete equation systems. Technical report, The Hauge:The Central Planning Bureau, The Netherlands.

USDA (1984). Interpretive guide for available water holding capacity of soils corrected for content of rock fragments and salts. *NRCS Missoula Conservation District. Montana.*

Uzoh, F. C. (2001). A height increment equation for ponderosa pine using precipitation and soil factors. *Forest Ecology and Management*, 142(1-3):193–203.

Uzoh, F. C. C., Dolph, K. L., and Anstead, J. R. (1998). Periodic Annual Diameter Increment After Overstory Removal in Mixed Conifer Stands. Research Paper PSW-RP-238, United States Department of Agriculture, Forest Service, Pacific Southwest Research Station, Berkeley, California.

Vanclay, J. K. (1994). *Modelling Forest Growth and Yield; Applications to Mixed Tropical Forests.* CAB International,, Oxon, UK.

Wagner, R. G. (1989). *Interspecific competition in young Douglas-fir plantations of the Oregon Coast Range.* PhD thesis, Oregon State University, Corvallis, Oregon.

Wagner, R. G. and Radosevich, S. R. (1991a). Interspecific competition and other factors influencing the performance of douglas-fir saplings in the oregon coast range. *Canadian Journal of Forest Research*, 21:829–835.

Wagner, R. G. and Radosevich, S. R. (1991b). Neighborhood predictors of interspecific competition in young douglas-fir plantations. *Canadian Journal of Forest Research*, 21:821–828.

Waring, R. H., Thies, W. G., and Muscato, D. (1980). Stem growth per unit of leaf area: A measure of tree vigor. *Forest Science*, 26(1):112–117.

Weller, D. E. (1989). The interspecific size-density relationship among crowded plant stands and its implications for the -3/2 power rule of self-thinning. *The American Naturalist*, 133(1):20–41.

Wensel, L. C., Daugherty, P. J., and Meerchaert, W. J. (1986). CACTOS User's Guide: The California Conifer Timber Output Simulator. Version 3.3. Agriculture Experiment Station Bulletin 1920, University of California Division of Agriculture and Natural Resources, Berkeley, California.

Wensel, L. C., Meerchaert, W. J., and Biging, G. S. (1987). Tree Height and Diameter Growth Models for Northern California Conifers. *Hilgardia*, 55(8).

Wensel, L. C. and Olsen, C. M. (1993). Tree Taper Models for Major Commercial California Conifers. Technical Report Research Note 33(revised), Northern California Forest Yield Cooperative, Department of Forestry and Resource Management, University of California, Berkeley, California.

West, G. B., Brown, J. H., and Enquist, B. J. (1997). A General Model for the Origin of Allometric Scaling Laws in Biology. *Nature*, 276:122–126.

West, P. W. (1980). Use of diameter increment and basal area increment in tree growth studies. *Canadian Journal of Forest Research*, 10(1):71–77.

Wykoff, W. R. (1990). A basal area increment model for individual conifers in the northern Rocky Mountains. *Forest Science*, 36(4):1077–1104.

Wykoff, W. R., Crookston, N. L., and Stage, A. R. (1982). User's Guide to the Stand Prognosis Model. Technical Report INT-133, USDA Forest Service.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zellner, A. and Theil, H. (1962). Three Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica*, 30(1):54–78.

Appendix

APPENDIX A

SYSTEMFIT DOCUMENTATION

Package 'systemfit'

December 26, 2002

Version 0.5-4

Date 2002/06/04

Title Simultaneous Equation Estimation Package

Author Jeff D. Hamann <jeff_hamann@hamanndonald.com>

Maintainer Jeff D. Hamann <jeff_hamann@hamanndonald.com>

Depends R (>= 1.3.0)

Description This package contains functions for fitting simultaneous systems of equations using Ordinary Least Sqaures (OLS), Two-Stage Least Squares (2SLS), and Three-Stage Least Squares (3SLS).

URL http://www.r-project.org, http://www.hamanndonald.com

R TOPICS DOCUMENTED:

`correlation` — *Correlation between Predictions from Equation i and j*

Description

`correlation` returns a vector of the correlations
between the preditions of two equations in a set of equations. The correlation between the predictions is defined as,
equation i and j and C_{ij} is the cross-equation variance-covariance matrix between equations i and j.

Usage

```
correlation.systemfit( results, eqni, eqnj )
```

Arguments

`results`	an object of type systemfit.system.
`eqni`	index for equation i
`eqnj`	index for equation j

Value

`correlation` returns a vector of the correlations between the predicted values in equation i and equation j.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Greene, W. H. (1993)

Econometric Analysis, Second Edition, Macmillan.

Hasenauer, H; Monserud, R and T. Gregoire. (1998)

Using Simultansous Regression Techniques with Individual-Tree Growth Models.

Forest Science. 44(1):87-95

Kmenta, J. (1997)

Elements of Econometrics, Second Edition, University of Michigan Publishing

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )
```

```
data( kmenta )
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
labels <- list( "demand", "supply" )
system <- list( demand, supply )

## perform 2SLS on each of the equations in the system
fit2sls <- twostage.systemfit( system, inst, labels, kmenta )
print( fit2sls )
print( varcov.systemfit( fit2sls ) )

## perform the 3SLS
fit3sls <- threestage.systemfit( system, inst, labels, kmenta )
print( fit3sls )
print( "covariance of residuals used for estimation (from 2sls)" )
print( varcov.systemfit( fit2sls ) )
print( "covariance of residuals" )
print( varcov.systemfit( fit3sls ) )
```

```
## examine the correlation between the predicted values

## of suppy and demand by plotting the correlation over

## the value of q

r12 <- correlation.systemfit( fit3sls, 1, 2 )

plot( q,
r12,
main="correlation between predictions from supply and demand" )


## examine the improvement of 3SLS over OLS by computing

## the ratio of the standard errors of the estimates

improve.ratio <- se.ratio.systemfit( fit2sls, fit3sls, 2 )

print( "summary values for the ratio in the std. err." )
print( "for the predictions" )

print( summary( improve.ratio ) )
```

`hausman` *Hausman's Test*

Description

`hausman` returns the Hausman's statistic for specification.

$$q'(V_1 - V_0)q$$

where V_1 and V_0 are the covb values from a `twostage` or `threestage` object and q is the difference in the b vectors from the `twostage` or `threestage` objects.

Usage

```
hausman.systemfit( results0, results1 )
```

Arguments

`results0`	the ith equation in the set of `twostage` or `threestage` objects
`results1`	the jth equation in the set of `twostage` or `threestage` objects

Value

`hausman.systemfit` returns the value of the test statistic.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Greene, W. H. (1993)
Econometric Analysis, Second Edition, Macmillan.
Hausman, J. A. (1978)
Specification Tests in Econometrics. *Economtrica.* 46:1251-1271.
Kmenta, J. (1997)
Elements of Econometrics, Second Edition, University of Michigan Publishing

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform the estimation and report the results for the whoel system

fit2sls <- twostage.systemfit( system, inst, labels, kmenta )

fit3sls <- threestage.systemfit( system, inst, labels, kmenta )


## perform the hausman test on the first equation

h <- hausman.systemfit( fit3sls[[1]], fit2sls[[1]] )

pval <- pchisq( h, dim( fit3sls[[1]]$covb )[1] )
```

`kmenta` *Partly Artificial Data on the U. S. Economy*

Description

These are partly contrived data from Kmenta (1986), constructed to illustrate estimation of a simultaneous-equation model.

The `kmenta` data frame has 20 rows and 5 columns.

Usage

```
data(kmenta)
```

Format

This data frame contains the following columns:

q food consumption per capita.

p ratio of food prices to general consumer prices.

d disposable income in constant dollars.

f ratio of preceding year's prices received by farmers to general consumer prices.

a time in years.

Details

The exogenous variables `d`, `f`, and `a` are based on real data; the endogenous variables `p` and `q` were generated by simulation.

Source

Kmenta, J. (1986)
Elements of Econometrics, Second Edition, Macmillan.

Examples

```
data(kmenta)
```

`ols` *Ordinary Least Squares Estimation*

Description

Fits a set of structural equations using Ordinary Least Squares. The resulting object is an array of fitting regression equations that contain information about the fitting process as well as the resulting parameter estimates, standard error estimates and covaraince matrix.

Usage

```
ols.systemfit( eqns, instruments, eqnlabels, data )
```

Arguments

`eqns`	a list of structural equations to be estimated; a regression constant is implied if not explicitly omitted.
`instruments`	one-sided model formula specifying instrumental variables.
`eqnlabels`	list of character vectors of names for the equation labels.
`data`	an optional data frame containing the variables in the model. By default the variables are taken from the environment from which twostage is called.

Value

`ols` returns a list of objects of class `ols`, with the following components:

- `n` number of observations.
- `p` number of parameters.
- `coefficients` parameter estimates.
- `V` estimated covariance matrix of coefficients.
- `s` residual standard error.
- `residuals` vector of residuals.
- `response` vector of response values.
- `X` model matrix.
- `Z` instrumental-variables matrix.
- `response.name` name of response variable, or expression evaluating to response.
- `formula` model formula.
- `instruments` one-sided formula for instrumental variables.
- `method` estimation method for the object, in this case, "OLS".
- `eqnlabel` the equation label from the labels list for the equation.
- `formula` model formula.
- `dfe` error degrees of freedom.
- `dfm` model degrees of freedom.
- `model.matrix` model matrix for the ith equation
- `model.frame` model frame for the ith equation
- `instruments` list of instruments for the set of equations.
- `response` y
- `predicted` predicted values
- `residuals` residuals
- `ztzinv` two-stage instrument regression matrix - ztzinv.
- `v` v
- `b` parameter estimates
- `n` number of observations for the ith equation
- `s` estimation of sigma
- `sse` sum of squares
- `mse` mean squared error
- `rmse` square root of mse or root mean squared error

`se`	estimates standard errors of the parameter estimates.
`t`	t values for b
`p`	p values for b
`r2`	r2
`adjr2`	adjusted r-squared
`covb`	covb

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Greene, W. H. (1993)
Econometric Analysis, Second Edition, Macmillan.
Kmenta, J. (1997)
Elements of Econometrics, Second Edition, University of Michigan Publishing

See Also

`twostage`, `threestage`

Examples

```
library( systemfit )



data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )
```

```
system <- list( demand, supply )


## perform OLS on each of the equations in the system
fit1sls <- ols.systemfit( system, inst, labels, kmenta )
print( fit1sls )
```

`print.systemfit.ols` *print.systemfit.ols*

Description

This function prints a summary of the system of equations.

Usage

```
print.systemfit.ols(x,digits=6,...)
```

Arguments

`x`	an object of type `ols.systemfit`.
`digits`	number of digits to print.
`...`	not used by user.

Value

`print.systemfit.ols` returns nothing.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform OLS on each of the equations in the system

fit1sls <- ols.systemfit( system, inst, labels, kmenta )


## print the results

print( fit1sls )
```

`print.systemfit.system` *print.systemfit.system*

Description

This function prints a summary of the system of equations.

Usage

```
print.systemfit.system(x,digits=6,...)
```

Arguments

`x`	an object of type `threestage.systemfit`.
`digits`	the number of digits to print.
`...`	not used by user.

Value

`print.systemfit.system` returns nothing.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )
```

```
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
labels <- list( "demand", "supply" )
system <- list( demand, supply )

## perform 2SLS on each of the equations in the system
fit3sls <- ols.systemfit( system, inst, labels, kmenta )

## print the results
print( fit3sls )
```

`print.systemfit.threestage` *print.systemfit.threestage*

Description

This function prints a summary of the system of equations.

Usage

```
print.systemfit.threestage(x,digits=6,...)
```

Arguments

`x`	an object of type `threestage.systemfit`.
`digits`	the number of digits to print.
`...`	not used by user.

Value

`print.systemfit.threestage` returns nothing.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )



data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform 2SLS on each of the equations in the system
```

```
fit3sls <- ols.systemfit( system, inst, labels, kmenta )

## print the results
print( fit3sls )
```

`print.systemfit.twostage` *print.systemfit.twostage*

Description

This function prints a summary of the system of equations.

Usage

```
print.systemfit.twostage(x,digits=6,...)
```

Arguments

`x`	an object of type `twostage.systemfit`.
`digits`	the number of digits to print.
`...`	not used by user.

Value

`print.systemfit.twostage` returns nothing.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
labels <- list( "demand", "supply" )
system <- list( demand, supply )

## perform 2SLS on each of the equations in the system
fit2sls <- twostage.systemfit( system, inst, labels, kmenta )

## print the results
print( fit2sls )
```

`se.ratio` *Ratio of the Standard Errors*

Description

`se.ratio.systemfit` returns a vector of the ratios of the standard errors of the predictions for two equations.

Usage

```
se.ratio.systemfit( resultsi, resultsj, eqni )
```

Arguments

`resultsi`	an object of type systemfit.system (`ols`, `twostage` or `threestage`.
`resultsj`	an object of type systemfit.system (`ols`, `twostage` or `threestage`.
`eqni`	index for equation to obtain the ratio of standard errors

Value

`se.ratio` returns a vector of the standard errors of the ratios for the predictions between the predicted values in equation i and equation j.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Hasenauer, H; Monserud, R and T. Gregoire. (1998)
Using Simultansous Regression Techniques with Individual-Tree Growth Models.
Forest Science. 44(1):87-95

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform 2SLS on each of the equations in the system

fit2sls <- twostage.systemfit( system, inst, labels, kmenta )

fit3sls <- threestage.systemfit( system, inst, labels, kmenta )


## print the results from the fits

print( fit2sls )

print( fit3sls )

print( "covariance of residuals used for estimation (from 2sls)" )

print( varcov.systemfit( fit2sls ) )

print( "covariance of residuals" )
```

```
print( varcov.systemfit( fit3sls ) )


## examine the correlation between the predicted values

## of suppy and demand by plotting the correlation over

## the value of q

r12 <- correlation.systemfit( fit3sls, 1, 2 )

plot( q,
r12,
main="correlation between predictions from supply and demand" )


## examine the improvement of 3SLS over OLS by computing

## the ratio of the standard errors of the estimates

improve.ratio <- se.ratio.systemfit( fit2sls, fit3sls, 2 )


print( "summary values for the ratio in the std. err." )
print( "for the predictions" )

print( summary( improve.ratio ) )
```

`summary.systemfit.ols` *summary.systemfit.ols*

Description

This function returns a summary of the system of equations.

Usage

```
summary.systemfit.ols(object,...)
```

Arguments

`object`	an object of type `ols.systemfit`.
`...`	not used by user.

Value

`summary.systemfit.ols` returns an object of type `systemfit.ols`.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )



data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )
```

```
## perform OLS on each of the equations in the system
fit1sls <- ols.systemfit( system, inst, labels, kmenta )

## print the results
print( fit1sls )
```

`summary.systemfit.system` *summary.systemfit.system*

Description

This function returns a summary of the system of equations.

Usage

```
summary.systemfit.system(object,...)
```

Arguments

`object`	an object of type `systemfit.system`.
`...`	not used by user.

Value

`summary.systemfit.system` returns an object of type `systemfit.system`.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

ols,twostage and threestage

Examples

```
library( systemfit )


data( kmenta )
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
labels <- list( "demand", "supply" )
system <- list( demand, supply )

## perform 3SLS on each of the equations in the system
fit3sls <- threestage.systemfit( system, inst, labels, kmenta )

## print the results
print( fit3sls )
```

summary.systemfit.threestage

summary.systemfit.threestage

Description

This function returns a summary of the system of equations.

Usage

```
summary.systemfit.threestage(object,...)
```

Arguments

`object`	an object of type `threestage.systemfit`.
`...`	not used by user.

Value

`summary.systemfit.threestage` returns an object of type `systemfit.threestage`.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )
```

```
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
labels <- list( "demand", "supply" )
system <- list( demand, supply )

## perform 3SLS on each of the equations in the system
fit3sls <- threestage.systemfit( system, inst, labels, kmenta )

## print the results
print( fit3sls )
```

---

summary.systemfit.twostage *summary.systemfit.twostage*

Description

This function returns a summary of the system of equations.

Usage

```
summary.systemfit.twostage(object,...)
```

Arguments

`object`	an object of type `twostage.systemfit`.
`...`	not used by user.

Value

`summary.systemfit.twostage` returns an object of type `systemfit.ols`.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform 2SLS on each of the equations in the system
```

```
fit2sls <- twostage.systemfit( system, inst, labels, kmenta )


## print the results

print( fit2sls )
```

---

threestage	*Three-Stage Least Squares Estimation*

Description

Fits a set of structural equations using Three-Stage Least Squares. The resulting object is an array of fitting regression equations that contain information about the fitting process as well as the resulting parameter estimates, standard error estimates and covaraince matrix.

Usage

```
threestage.systemfit( eqns, instruments, eqnlabels, data )
```

Arguments

`eqns`	a list of structural equations to be estimated; a regression constant is implied if not explicitly omitted.
`instruments`	one-sided model formula specifying instrumental variables.
`data`	an optional data frame containing the variables in the model. By default the variables are taken from the environment from which twostage is called.
`eqnlabels`	list of character vectors of names for the equation labels.

Value

threestage returns a list of objects of class threestage, with the following components:

- n: number of observations.
- p: number of parameters.
- coefficients: parameter estimates.
- V: estimated covariance matrix of coefficients.
- s: residual standard error.
- residuals: vector of residuals.
- response: vector of response values.
- X: model matrix.
- Z: instrumental-variables matrix.
- response.name: name of response variable, or expression evaluating to response.
- formula: model formula.
- instruments: one-sided formula for instrumental variables.
- method: estimation method for the object, in this case, "3SLS".
- eqnlabel: the equation label from the labels list for the equation.
- formula: model formula.
- dfe: error degrees of freedom.
- dfm: model degrees of freedom.
- model.matrix: model matrix for the ith equation
- model.frame: model frame for the ith equation
- instruments: list of instruments for the set of equations.
- response: y
- predicted: predicted values
- residuals: residuals
- ztzinv: two-stage instrument regression matrix - ztzinv.
- v: v
- b: parameter estimates
- n: number of observations for the ith equation
- s: estimation of sigma
- sse: sum of squares
- mse: mean squared error

`rmse`	square root of mse or root mean squared error
`se`	estimates standard errors of the parameter estimates.
`t`	t values for b
`p`	p values for b
`r2`	r2
`adjr2`	adjusted r-squared
`covb`	covb

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Greene, W. H. (1993)
Econometric Analysis, Second Edition, Macmillan.
Kmenta, J. (1997)
Elements of Econometrics, Second Edition, University of Michigan Publishing

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )

data( kmenta )
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
```

```
labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform 2SLS on each of the equations in the system

fit2sls <- twostage.systemfit( system, inst, labels, kmenta )

print( fit2sls )

print( varcov.systemfit( fit2sls ) )


fit3sls <- threestage.systemfit( system, inst, labels, kmenta )

print( fit3sls )

print( "covariance of residuals used for estimation (from 2sls)" )

print( varcov.systemfit( fit2sls ) )

print( "covariance of residuals" )

print( varcov.systemfit( fit3sls ) )


## examine the correlation between the predicted values

## of suppy and demand by plotting the correlation over

## the value of q

r12 <- correlation.systemfit( fit3sls, 1, 2 )

plot( q,
r12,
main="correlation between predictions from supply and demand" )
```

```
## examine the improvement of 3SLS over OLS by computing

## the ratio of the standard errors of the estimates

improve.ratio <- se.ratio.systemfit( fit2sls, fit3sls, 2 )

print( "summary values for the ratio in the std. err." )
print( "for the predictions" )

print( summary( improve.ratio ) )


## perform the hausman test

h <- hausman.systemfit( fit3sls[[1]], fit2sls[[1]] )

pval <- pchisq( h, dim( fit3sls[[1]]$covb )[1] )
```

`threestage.cov` *Variance-Covariance Matrix*

Description

This function returns a variance-covariance estimation matrix from a resulting simultaneous estimation object such as type `threestage`.

Usage

```
threestage.cov( results, eqni, eqnj )
```

Arguments

`results`	a set of 3SLS objects returned from `threestage`
`eqni`	the ith equation in the set of `threestage` objects
`eqnj`	the jth equation in the set of `threestage` objects

Value

`threestage.cov` returns a submatrix from the variance-covariance matrix from the variance-covariance matrix used for estimation during 3SLS.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Hasenauer, H; Monserud, R and T. Gregoire. (1998)
Using Simultansous Regression Techniques with Individual-Tree Growth Models.
Forest Science. 44(1):87-95

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a
```

```
inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform the estimation and report the results for the whoel system

fit3sls <- threestage.systemfit( system, inst, labels, kmenta )

print( fit3sls )


## get the variance-covariance matrix used for estimation

print( "covariance of residuals used for estimation (from 2sls)" )

print( threestage.cov( fit3sls, 1, 2 ) )
```

`twostage` *Two-Stage Least Squares Estimation*

Description

Fits a set of structural equations using Two-Stage Least Squares. The resulting object is an array of fitting regression equations that contain information about the fitting process as well as the resulting parameter estimates, standard error estimates and covarAince matrix.

Usage

```
twostage.systemfit( eqns, instruments, eqnlabels, data )
```

Arguments

`eqns`	a list of structural equations to be estimated; a regression constant is implied if not explicitly omitted.
`instruments`	one-sided model formula specifying instrumental variables.
`data`	an optional data frame containing the variables in the model. By default the variables are taken from the environment from which twostage is called.
`eqnlabels`	list of character vectors of names for the equation labels.

Value

`twostage` returns a list of objects of class `twostage`, with the following components:

`n`	number of observations.
`p`	number of parameters.
`coefficients`	parameter estimates.
`V`	estimated covariance matrix of coefficients.
`s`	residual standard error.
`residuals`	vector of residuals.
`response`	vector of response values.
`X`	model matrix.
`Z`	instrumental-variables matrix.
`response.name`	name of response variable, or expression evaluating to response.
`formula`	model formula.
`instruments`	one-sided formula for instrumental variables.
`method`	estimation method for the object, in this case, "2SLS".
`eqnlabel`	the equation label from the labels list for the equation.
`formula`	model formula.
`dfe`	error degrees of freedom.
`dfm`	model degrees of freedom.
`model.matrix`	model matrix for the ith equation
`model.frame`	model frame for the ith equation

`instruments`	list of instruments for the set of equations.
`response`	y
`predicted`	predicted values
`residuals`	residuals
`ztzinv`	two-stage instrument regression matrix - ztzinv.
`v`	v
`b`	parameter estimates
`n`	number of observations for the ith equation
`s`	estimation of sigma
`sse`	sum of squares
`mse`	mean squared error
`rmse`	square root of mse or root mean squared error
`se`	estimates standard errors of the parameter estimates.
`t`	t values for b
`p`	p values for b
`r2`	r2
`adjr2`	adjusted r-squared
`covb`	covb

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Greene, W. H. (1993)
Econometric Analysis, Second Edition, Macmillan.
Kmenta, J. (1997)
Elements of Econometrics, Second Edition, University of Michigan Publishing

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )
attach( kmenta )
demand <- q ~ p + d
supply <- q ~ p + f + a
inst <- ~ d + f + a
labels <- list( "demand", "supply" )
system <- list( demand, supply )

## perform 2SLS on each of the equations in the system
fit2sls <- twostage.systemfit( system, inst, labels, kmenta )
fit3sls <- threestage.systemfit( system, inst, labels, kmenta )

## print the results
print( fit2sls )
print( fit3sls )
print( "covariance of residuals used for estimation (from 2sls)" )
print( varcov.systemfit( fit2sls ) )
print( "covariance of residuals" )
```

```
print( varcov.systemfit( fit3sls ) )


## examine the correlation between the predicted values

## of suppy and demand by plotting the correlation over

## the value of q

r12 <- correlation.systemfit( fit3sls, 1, 2 )

plot( q,
r12,
main="correlation between predictions from supply and demand" )


## examine the improvement of 3SLS over OLS by computing

## the ratio of the standard errors of the estimates

improve.ratio <- se.ratio.systemfit( fit2sls, fit3sls, 2 )

print( "summary values for the ratio in the std. err." )
print( "for the predictions" )

print( summary( improve.ratio ) )
```

---

varcov *Variance-Covariance*

Description

The function returns the variance-covaraince of the residuals for a set of equations from the residuals. The values of the elements are defined as,

$$\frac{e_i' e_j}{\sqrt{n_i n_j}}$$

where e_i and e_j are the residuals and n_i and n_j are the error degrees of freedom for equations i and j.

Usage

```
varcov.systemfit( results )
```

Arguments

`results`	an object of type `twostage.systemfit`.

Value

`varcov` returns a variance-covaraince matrix of the residuals from a set of objects of class `twostage`.

Author(s)

Jeff D. Hamann ⟨jeff_hamann@hamanndonald.com⟩

References

Greene, W. H. (1993)
Econometric Analysis, Second Edition, Macmillan.
Kmenta, J. (1997)
Elements of Econometrics, Second Edition, University of Michigan Publishing

See Also

`ols`,`twostage` and `threestage`

Examples

```
library( systemfit )


data( kmenta )

attach( kmenta )

demand <- q ~ p + d

supply <- q ~ p + f + a

inst <- ~ d + f + a

labels <- list( "demand", "supply" )

system <- list( demand, supply )


## perform 2SLS on each of the equations in the system

fit2sls <- twostage.systemfit( system, inst, labels, kmenta )

fit3sls <- threestage.systemfit( system, inst, labels, kmenta )


## print the results

print( fit2sls )

print( fit3sls )

print( "covariance of residuals used for estimation (from 2sls)" )

print( varcov.systemfit( fit2sls ) )

print( "covariance of residuals" )
```

```
print( varcov.systemfit( fit3sls ) )
```

APPENDIX B

SYSTEMFIT SOURCE CODE

```
##  $Id: systemfit.R,v 1.3 2002/11/19 08:36:38 hamannj Exp $


## performs two-stage least squares on the system of equations
ols.systemfit <- function(
                          eqns,
                          instruments,
                          eqnlabels,
                          data )
{
  results <- list()
  resulti <- list()

  for(i in 1:length( eqns ) )
    {

      ## perform the two stage least squares regression
      y   <- eval( attr( terms( eqns[[i]] ), "variables" )[[2]] )
      x   <- model.matrix( eqns[[i]] )

      v <- diag( dim( model.matrix( eqns[[i]] )[2] ) )
      ## two stage least squares results...
      b   <- solve( t(x) %*% x ) %*% t(x) %*% y
      resids   <- y - x %*% b
      n   <- length( y )
      p   <- ncol( x )
      s   <- sum(resids^2)/(n - p)
      dfe   <- n - p
      se   <- sqrt( diag( solve( t(x) %*% x ) ) * s )
      t   <- b/se
      prob   <- 2.0*(1.0 - pt(abs(t), dfe))
      mse   <- ( s * dfe / dfe )
      rmse   <- sqrt( mse )
      r2   <- 1.0 - ((t(resids)%*%resids)/(t(y)%*%y-n*mean(y)^2))
      adjr2   <- 1.0 - ((n-1)/(n-p))*(1.0-r2)
      covb   <- solve( t(x) %*% x )

      ## build the "return" structure for the 2sls part
      resulti$method      <- "ols"
      resulti$eqnlabel     <- eqnlabels[[i]]
      resulti$formula      <- eqns[[i]]
      resulti$dfe      <- dfe
      resulti$dfm      <- n - dfe
      resulti$model.matrix <- model.matrix(eqns[[i]] )
      resulti$model.frame  <- model.frame(eqns[[i]] )
      resulti$instruments  <- inst
      resulti$response     <- y
```

```
      resulti$predicted    <- x %*% b
      resulti$residuals    <- resids
      ##resulti$ztzinv      <- ztzinv
      resulti$v      <- v
      resulti$b      <- b
      names(resulti$b)     <- colnames( model.matrix( eqns[[i]] ) )
      resulti$n      <- n
      resulti$s      <- s
      resulti$sse      <- s * dfe
      resulti$mse      <- mse
      resulti$rmse      <- rmse
      resulti$se      <- se
      resulti$t      <- t
      resulti$p      <- prob
      resulti$r2      <- r2
      resulti$adjr2      <- adjr2
      resulti$covb      <- covb
      class(resulti)      <- "systemfit.ols"
      results[[i]]      <- resulti

    }

  class(results) <- "systemfit.system"
  ols <- results

}


## this function produces a table for a single equation
## in a system of equations
summary.systemfit.ols <- function(object,...)
{
  summary.systemfit.ols <- object
  summary.systemfit.ols
}

## now print the object that comes from the fits...
print.systemfit.ols <- function( x, digits=6, ... )
{

  object <- x

  save.digits <- unlist(options(digits=digits))
  on.exit(options(digits=save.digits))

  cat("\n")
  cat( paste( attr( object, "class" ),
             "estimates for", object$eqnlabel, "\n" ) )

  cat("Model Formula: ")
  print(object$formula)
  cat("Instruments: ")
  print(object$instruments)
```

```
  cat("\n")

  Signif <- symnum(object$p, corr = FALSE, na = FALSE,
                   cutpoints = c(0,  .001,.01,.05, .1, 1),
                   symbols   =  c("***","**","*","."," "))

  table <- cbind(round( object$b, digits ),
                 round( object$se, digits ),
                 round( object$t, digits ),
                 round( object$p, digits ),
                 Signif)

  rownames(table) <- names(object$b)
  colnames(table) <- c("Estimate",
                       "Std. Error","t value","Pr(>|t|)","")

  print.matrix(table, quote = FALSE, right = TRUE )
  cat("---\nSignif. codes: ",attr(Signif,"legend"),"\n")

  cat(paste("\nResidual standard error:", round(object$s, digits),
          "on", object$dfe, "degrees of freedom\n"))

  cat( paste( "DF-Error:", round(object$dfe, digits),
             "DF-Model:", round(object$dfm, digits),
             "\n" ) )

  cat( paste( "SSE:", round(object$sse, digits),
             "MSE:", round(object$s, digits),
             "Root MSE:", round( sqrt(object$s), digits), "\n" ) )

  cat( paste( "Multiple R-Squared:", round(object$r2, digits),
             "Adjusted R-Squared:", round(object$adjr2, digits),
             "\n" ) )
  cat("\n")

}


## performs two-stage least squares on the system of equations
twostage.systemfit <- function(
                              eqns,
                              instruments,
                              eqnlabels,
                              data )
{

  results <- list()
  resulti <- list()

  for(i in 1:length( eqns ) )
    {

      ## perform the two stage least squares regression
```

```
y   <- eval( attr( terms( eqns[[i]] ), "variables" )[[2]] )
x   <- model.matrix( eqns[[i]] )
z   <- model.matrix( inst )
ztzinv   <- solve( t(z) %*% z )
v   <- solve( t(x) %*% z %*% ztzinv %*% t(z) %*% x )

## two stage least squares results...
b   <- v %*% t(x) %*% z %*% ztzinv %*% t(z) %*% y
resids   <- y - x %*% b
n   <- length( y )
p   <- ncol( x )
s   <- sum(resids^2)/(n - p)
dfe   <- n - p
se   <- sqrt(diag(s*v))
t   <- b/se
prob   <- 2.0*(1.0 - pt(abs(t), dfe))
mse   <- ( s * dfe / dfe )
rmse   <- sqrt( mse )
r2   <- 1.0 - ((t(resids)%*%resids)/(t(y)%*%y-n*mean(y)^2))
adjr2   <- 1.0 - ((n-1)/(n-p))*(1.0-r2)
covb   <- v * s

## get the residuals from the 2sls on the instruments
instres <- lsfit( model.frame( inst ),
                  model.matrix( eqns[[i]] ) )$coef
temp2 <- model.matrix( inst ) %*% instres
resulti$instres <- instres
resulti$tslsres <- temp2

## build the "return" structure for the 2sls part
resulti$method      <- "2sls"
resulti$eqnlabel     <- eqnlabels[[i]]
resulti$formula      <- eqns[[i]]
resulti$dfe      <- dfe
resulti$dfm      <- n - dfe
resulti$model.matrix <- model.matrix(eqns[[i]] )
resulti$model.frame  <- model.frame(eqns[[i]] )
resulti$instruments  <- inst
resulti$response     <- y
resulti$predicted    <- x %*% b
resulti$residuals    <- resids
resulti$ztzinv      <- ztzinv
resulti$v      <- v
resulti$b      <- b
names(resulti$b)     <- colnames( model.matrix( eqns[[i]] ) )
resulti$n      <- n
resulti$s      <- s
resulti$sse      <- s * dfe
resulti$mse      <- mse
resulti$rmse      <- rmse
resulti$se      <- se
resulti$t      <- t
resulti$p      <-  prob
```

```
      resulti$r2      <- r2
      resulti$adjr2      <- adjr2
      resulti$covb     <- covb
      class(resulti)     <- "systemfit.twostage"
      results[[i]]     <- resulti

    }

  class(results) <- "systemfit.system"
  twostage <- results
}


summary.systemfit.twostage <- function(object,...)
{
  summary.systemfit.twostage <- object
  summary.systemfit.twostage
}

## now print the object that comes from the fits...
print.systemfit.twostage <- function( x,digits=6,... )
{

  object <- x

  save.digits <- unlist(options(digits=digits))
  on.exit(options(digits=save.digits))

  cat("\n")
  cat( paste( attr( object, "class" ),
           "estimates for", object$eqnlabel, "\n" ) )

  cat("Model Formula: ")
  print(object$formula)
  cat("Instruments: ")
  print(object$instruments)
  cat("\n")

  Signif <- symnum(object$p, corr = FALSE, na = FALSE,
                   cutpoints = c(0,  .001,.01,.05, .1, 1),
                   symbols   =  c("***","**","*","."," "))

  table <- cbind(round( object$b, digits ),
                 round( object$se, digits ),
                 round( object$t, digits ),
                 round( object$p, digits ),
                 Signif)

  rownames(table) <- names(object$b)
  colnames(table) <- c("Estimate",
                       "Std. Error","t value","Pr(>|t|)","")

  print.matrix(table, quote = FALSE, right = TRUE )
```

```
  cat("---\nSignif. codes: ",attr(Signif,"legend"),"\n")

  cat(paste("\nResidual standard error:", round(object$s, digits),
          "on", object$dfe, "degrees of freedom\n"))

  cat( paste( "DF-Error:", round(object$dfe, digits),
           "DF-Model:", round(object$dfm, digits),
           "\n" ) )

  cat( paste( "SSE:", round(object$sse, digits),
           "MSE:", round(object$s, digits),
           "Root MSE:", round( sqrt(object$s), digits), "\n" ) )

  cat( paste( "Multiple R-Squared:", round(object$r2, digits),
           "Adjusted R-Squared:", round(object$adjr2, digits),
           "\n" ) )
  cat("\n")

}

## performs two-stage least squares on the system of equations
threestage.systemfit <- function(
                                 eqns,
                                 instruments,
                                 eqnlabels,
                                 data )
{

  results <- list()
  resulti <- list()
  u2 <-  matrix( 0, dim(data)[1], length( eqns ) )

  ## perform the two-stage fits
  tsls <- twostage.systemfit(
                             eqns,
                             instruments,
                             eqnlabels,
                             data )


  ## these are the ones that wil be used to build the big matrix
  t3 <- NULL
  bigb <- NULL
  bigy <- NULL
  bigt <- NULL
  bigse <- NULL
  bigp <- NULL

  for(i in 1:length( eqns ) )
    {
      ## build the final large matrix...
      tr <- NULL

      ## get the dimensions of the current matrix
      for(j in 1:length( eqns ) )
```

```
      {
        if( i == j )
          {
            tr <- cbind( tr, tsls[[i]]$tslsres )
          }
        else
          {
            ## bind the zero matrix to the row
            di <- dim( model.matrix( eqns[[j]] ) )[1]
            dj <- dim( model.matrix( eqns[[j]] ) )[2]
            tr <- cbind( tr, matrix( 0, di, dj ) )
          }
      }

    t3 <- rbind( t3, tr )

    ## now add the rows to the bigX matrix
    ## or should this be the new fitted y values
    ## from the two stage least squares fits...
    y      <- eval( attr( terms( eqns[[i]] ), "variables" )[[2]] )
    bigy <- rbind( bigy, matrix( y ) )
  }

## get the variance-covariance matrix from the two stage results
varcov <- varcov.systemfit( tsls )

parta <- kronecker( solve( varcov ),
                   diag( dim( model.matrix( eqns[[1]] ) )[1] ) )
part1 <- solve( t(t3) %*% parta %*% t3 ) # covariance matrix
part2 <- t(t3) %*% parta %*% bigy
bigb <- part1 %*% part2

## compute the se, t, and p values...
bigse <- matrix( sqrt( diag( part1 ) ) )
bigt <- bigb/bigse

## extract the results
idx <- matrix( 0, length( eqns ), 2 )
for(i in 1:length( eqns ) )
  {


    ## get the index for stripping out the estimates
    if( i == 1 )
      {
        idx[i,1] <- 1
        idx[i,2] <- dim( model.matrix( eqns[[i]] ) )[2]
      }
    else
      {
        idx[i,1] <- idx[i-1,2]+1
        idx[i,2] <- idx[i,1] +
          dim( model.matrix(eqns[[i]]))[2]-1
```

```
  }

start1 <- idx[i,1]
start2 <- idx[i,2]

## tree stage least squares results...
x   <- model.matrix( eqns[[i]] )
y   <- eval( attr( terms( eqns[[i]] ), "variables" )[[2]] )
b   <- matrix( bigb[start1:start2] )
resids   <- y - x %*% b
n   <- length( y )
p   <- ncol( x )
s   <- sum(resids^2)/(n - p)
dfe   <- n - p
se   <- matrix( bigse[start1:start2] )
t   <- matrix( bigt[start1:start2] )
prob   <- 2.0*(1.0 - pt(abs(t), dfe))
mse   <- ( s * dfe / dfe )
rmse   <- sqrt( mse )
r2   <- 1.0 - ((t(resids)%*%resids)/(t(y)%*%y-n*mean(y)^2))
adjr2   <- 1.0 - ((n-1)/(n-p))*(1.0-r2)

## get the parameter var-cov matrix fo the eq
icol   <- ncol( model.matrix( eqns[[i]] ) )
jcol   <- ncol( model.matrix( eqns[[i]] ) )
startrow  <- idx[i,1]
endrow   <- idx[i,2]
startcol  <- idx[i,1]
endcol   <- idx[i,2]
covb <- matrix(
          part1[startrow:endrow,startcol:endcol], icol, jcol )

## build the "return" structure for the 3sls part
resulti$method       <- "3sls"
resulti$eqnlabel      <- eqnlabels[[i]]
resulti$formula       <- eqns[[i]]
resulti$dfe      <- dfe
resulti$dfm      <- n - dfe
resulti$model.matrix <- model.matrix(eqns[[i]] )
resulti$model.frame  <- model.frame(eqns[[i]] )
resulti$instruments  <- inst
resulti$response      <- tsls[[i]]$reponse
resulti$predicted     <- x %*% b
resulti$residuals     <- resids
resulti$ztzinv       <- tsls[[i]]$ztzinv
resulti$v      <- tsls[[i]]$v
resulti$b      <- b
names(resulti$b)      <- colnames( model.matrix( eqns[[i]] ) )
resulti$n      <- n
resulti$s      <- s
resulti$sse      <- s * dfe
resulti$mse      <- mse
```

```
      resulti$rmse      <- rmse
      resulti$se      <- se
      resulti$t      <- t
      resulti$p      <- prob
      resulti$r2      <- r2
      resulti$adjr2      <- adjr2
      resulti$systemcovb   <- part1
      resulti$covb      <- covb
      class(resulti)      <- "systemfit.threestage"
      results[[i]]      <- resulti

    }

  class(results) <- "systemfit.system"
  threestage <- results
}

summary.systemfit.threestage <- function(object,...)
{
  summary.systemfit.threestage <- object
  summary.systemfit.threestage
}

## now print the object that comes from the fits...
print.systemfit.threestage <- function( x,digits=6,... )
{

  object <- x

  save.digits <- unlist(options(digits=digits))
  on.exit(options(digits=save.digits))

  cat("\n")
  cat( paste( attr( object, "class" ),
             "estimates for", object$eqnlabel, "\n" ) )

  cat("Model Formula: ")
  print(object$formula)
  cat("Instruments: ")
  print(object$instruments)
  cat("\n")

  Signif <- symnum(object$p, corr = FALSE, na = FALSE,
                   cutpoints = c(0,  .001,.01,.05, .1, 1),
                   symbols   =  c("***","**","*","."," "))

  table <- cbind(round( object$b, digits ),
                 round( object$se, digits ),
                 round( object$t, digits ),
                 round( object$p, digits ),
                 Signif)

  rownames(table) <- names(object$b)
```

```
  colnames(table) <- c("Estimate",
                       "Std. Error","t value","Pr(>|t|)","")

  print.matrix(table, quote = FALSE, right = TRUE )
  cat("---\nSignif. codes: ",attr(Signif,"legend"),"\n")

  cat(paste("\nResidual standard error:", round(object$s, digits),
            "on", object$dfe, "degrees of freedom\n"))

  cat( paste( "DF-Error:", round(object$dfe, digits),
              "DF-Model:", round(object$dfm, digits),
              "\n" ) )

  cat( paste( "SSE:", round(object$sse, digits),
              "MSE:", round(object$s, digits),
              "Root MSE:", round( sqrt(object$s), digits), "\n" ) )

  cat( paste( "Multiple R-Squared:", round(object$r2, digits),
              "Adjusted R-Squared:", round(object$adjr2, digits),
              "\n" ) )
  cat("\n")

}


## this function returns the variance-covariance matrix
## from the results set for equation ij
threestage.cov <- function( results, eqni, eqnj )
{

  ## get the information about eqni and enqj
  ## get the size of the array for the matrix
  ## you are going to extract
  icol    <- ncol( results[[eqni]]$model.matrix )
  jcol    <- ncol( results[[eqnj]]$model.matrix )

  ## now get the offsets
  ## 1 - start row
  ## 2 - end row

  rows <- matrix( 0, length( results ), 2 )
  ##cols <- matrix( 0, length( results ), 2 )
  for(i in 1:length( results ) )
    {
      ## get the index for stripping out the estimates
      if( i == 1 )
        {
          rows[i,1] <- 1
          rows[i,2] <- dim( results[[i]]$model.matrix)[2]
        }
      else
        {
          rows[i,1] <- rows[i-1,2]+1
          rows[i,2] <- rows[i,1] + dim( results[[i]]$model.matrix )[2]-1
```

```
        }
    }
  startrow <- rows[eqni,1]
  endrow  <- rows[eqni,2]

  startcol <- rows[eqnj,1]
  endcol  <- rows[eqnj,2]

  test <- matrix(
          results[[1]]$systemcovb[startrow:endrow,startcol:endcol],
                 icol, jcol )

}

## this function returns test statistic for
## the hausman test which.... i forget, but people want to see it...
## from the sas docs
## given 2 estimators, b0 abd b1, where under the null hypothesis,
## both are consistent, but only b0 is asympt. efficient and
## under the alter. hypo only b1 is consistent, so the statistic (m) is
hausman.systemfit <- function( results0, results1 )
{

  v0 <- results0$covb
  v1 <- results1$covb
  q  <- results1$b - results0$b

  hausman <- t( q ) %*% ( v1 - v0 ) %*% q

}


## this function returns the covariance of the residuals
## the method will return the same matrix values as are
## returned in SAS in proc model
varcov.systemfit <- function( results )
{

  u2 <-  matrix( 0, length( results ), length( results ) )

  ## use bind to create a vector for the residuals
  for(i in 1:length( results ) )
    {
      ## use bind to create a vector for the residuals
      for(j in 1:length( results ) )
        {
          ri <- results[[i]]$residuals
          dfei <- results[[i]]$dfe

          rj <- results[[j]]$residuals
          dfej <- results[[j]]$dfe

          ## from SAS
```

```
          cvij <- ( t( ri ) %*% rj ) / ( sqrt( dfei * dfej ) )

          u2[i,j] <- cvij
        }

    }

  varcov <- u2
  varcov
}

## this function returns a vector of the
## cross-equation corrlations between eq i and eq j
## from the results set for equation ij
correlation.systemfit <- function( results, eqni, eqnj )
{

  cij <- threestage.cov( results, eqni, eqnj )
  cii <- threestage.cov( results, eqni, eqni )
  cjj <- threestage.cov( results, eqnj, eqnj )

  rij <- NULL

  for(i in 1:results[[1]]$n )
    {

      xik <- model.matrix( results[[eqni]]$formula )[i,]
      xjk <- model.matrix( results[[eqnj]]$formula )[i,]

      top <- xik %*% cij %*% xjk
      bottom <- sqrt( ( xik %*% cii %*% xik ) *
                      ( xjk %*% cjj %*% xjk ) )
      rijk <- top / bottom

      rij <- rbind( rij, rijk )

    }

  correlation <- rij
  correlation
}


## this function returns a vector of the
## cross-equation corrlations between eq i and eq j
## from the results set for equation ij
## you need to put some check in here to make sure both
## are the name type

## determines the improvement of resultsj (3sls) over
## resultsi (2sls) for equation i and returns a matrix
## of the values, so you can examine the range, mean, etc
se.ratio.systemfit <- function( resultsi, resultsj, eqni )
{
```

```
  ratio <- NULL

  for(i in 1:resultsi[[1]]$n )
    {

      xik <- model.matrix( resultsi[[eqni]]$formula )[i,]

      top    <- sqrt( xik %*% resultsi[[eqni]]$covb %*% xik )
      bottom <- sqrt( xik %*% resultsj[[eqni]]$covb %*% xik )
      rk     <- top / bottom

      ratio <- rbind( ratio, rk )

    }

  se.ratio <- ratio
  se.ratio
}

summary.systemfit.system <- function(object,...)
{
  summary.systemfit.system <- object
  summary.systemfit.system
}


## now print the object that comes from the fits...
print.systemfit.system <- function( x,digits=6,... )
{

  object <- x

  save.digits <- unlist(options(digits=digits))
  on.exit(options(digits=save.digits))


  table <- NULL
  labels <- NULL

  cat("\n")
  cat("systemfit results: \n")
  cat("\n")

  for(i in 1:length( object ) )
    {
      row <- NULL
      row <- cbind(
                   round( object[[i]]$dfm, digits ),
                   round( object[[i]]$dfe, digits ),
                   round( object[[i]]$sse, digits ),
                   round( object[[i]]$mse, digits ),
                   round( object[[i]]$rmse, digits ),
                   round( object[[i]]$r2, digits ),
                   round( object[[i]]$adjr2, digits ) )
```

```
      table <- rbind( table, row )
      labels <- rbind( labels, object[[i]]$eqnlabel )
    }

  rownames(table) <- c( labels )
  colnames(table) <- c(
                       "DF Model",
                       "DF Error",
                       "SSE",
                       "MSE",
                       "RMSE",
                       "R2",
                       "Adj R2" )

  print.matrix(table, quote = FALSE, right = TRUE )
  cat("\n")

  cat("The variance-covariance matrix\n")
  vc <- varcov.systemfit( object )
  rownames(vc) <- labels
  colnames(vc) <- labels
  print( vc )


  ## now print the individual equations
  for(i in 1:length( object ) )
    {
      print( object[[i]], digits )
    }

  save.digits <- unlist(options(digits=digits))
  on.exit(options(digits=save.digits))

}


## this function returns the covariance of the residuals
## the method will return the same matrix values as are
## returned in SAS in proc model
cor.systemfit <- function( results )
{

  ##u2 <-  matrix( 0, length( results ), length( results ) )

  ## use bind to create a vector for the residuals
  for(i in 1:length( results ) )
    {
      cm <- cbind( cm, results[[i]]$residuals )
    }

  cor <- cor( cm )
  cor
}
```