ECONOMETRIC MODELING OF SWEETGUM STEM BIOMASS USING THE IML AND SYSLIN PROCEDURES

Bernard R. Parresol, Southern Forest Experiment Station
Charles E. Thomas, Southern Forest Experiment Station

ABSTRACT

A recent development in tree-biomass estimation procedures has been the introduction of the density-integral methodology. This technique was developed on the basis of some simple calculus of mass theory. In the initial development of the technique, sectional bole weight was derived from the integral equation formed from a stem-profile function and a specific gravity function fitted independently. Recognizing the contemporaneous correlations between the equations, and the need to incorporate observed stem-weight in the fitting process, a more efficient parameter estimation approach is available through econometric methods. In this paper a simultaneous system of two equations is derived for sweetgum (Liquidambar styraciflua L.) and parameters are estimated using joint-generalized least squares (also called seemingly unrelated regressions or SUR) with cross-equation constraints. The matrix algebra necessary to fit the model and the corresponding SAS® software code used in PROC IML and PROC SYSLIN are shown. A test for contemporaneous correlation showed highly significant correlation between equations. Parameter estimates from SUR estimation had smaller standard errors than from ordinary least squares estimation.

INTRODUCTION

As early as 1950, the use of weight as a measure of wood quantity has been practiced by many of the larger companies in North America and Northern Europe. The use of weight measurement has steadily grown in importance with the increased demand for wood fiber and fuel. This paper refines a new approach for stem biomass estimation that was introduced by Parresol and Thomas (1989). This new approach uses a generalized integral model for the prediction of dry-weight yield of wood for any portion of a tree bole, given diameter at breast height (DBH) and total tree height (H).

DENSITY-INTEGRAL MODEL

The weight of an object is simply its volume multiplied by its density. Parresol and Thomas (1989) in deriving a tree bole density-integral model utilized some simple calculus of mass theory. Many standard calculus texts define mass (M) for a lamina with a continuous, nonconstant density function \( p(x,y) \) as

\[ M = \int \int_R p(x,y) \, dA \]

The 3-dimensional structure of a tree is easily conceptualized as a 2-dimensional lamina by taking cross-sectional area \( (y) \) as one dimension and height \( (h) \) as the other dimension (Figure 1). As we move from lower limit \( h_l \) to upper limit \( h_u \) in Figure 1, the variable \( y \) is seen to vary from 0 to upper limit \( f(h) \). This establishes the following model for stem biomass:

\[ w = \int_{h_l}^{h_u} \int_0^{f(h)} p(h,y) \, dy \, dh + \epsilon \]

where \( f(h) \) is an equation expressing taper in cross-sectional area as a function of height, \( w \) is bole dry weight of wood between limits \( h_l \) and \( h_u \), and \( \epsilon \) is residual error. Typically stem profile (cross-sectional area) is modeled using relative height \( (x = h/H) \) instead of actual height. Performing a change of variable from \( h \) to \( x \) results in the following generalized stem biomass model:

\[ w = H \int_x^{h_u} f(x) \int_0^{f(x)} p(x,y) \, dy \, dx + \epsilon \]  \hspace{1cm} (1) \]

For a specific biomass model one needs to define \( p \) and \( f \).

JOINT-GENERALIZED LEAST SQUARES FOR EFFICIENT ESTIMATES

In Parresol and Thomas (1989), working with slash pine (Pinus elliotti Engelm. var. elliottii), models for \( p(x,y) \) and \( f(x) \) were fitted using ordinary least squares (OLS) to obtain parameter estimates which were then substituted into Equation (1) for prediction of bole weight. This could lead to inefficient estimates of the parameters. One would expect weight, density, and volume to be correlated at the same measurement bolt on the tree. This correlation can be termed 'contemporaneous' to be consistent with econometric methods terminology. Additionally, observed stem weight needs to be incorporated in the fitting process. Joint-generalized least squares, more commonly known as seemingly unrelated regressions (SUR), takes into account contemporaneous correlations and leads to efficient estimates. Other econometric techniques, such as two-
and three-stage least squares (2SLS and 3SLS) are appropriate and indeed necessary for efficient and/or consistent estimates when dependent (endogenous) variables occur on the right-hand sides (RHS) of systems of equations.

**MATERIAL AND METHODS**

**Data**

The data base consists of 45 felled sweetgum (*Liquidambar styraciflua* L.) trees from natural stands. Trees were cut at varying stump heights, from 0.5 ft to 1.5 ft. Diameter inside bark in inches and specific gravity were measured at 5-ft intervals along the stem. Total tree height (H) in feet and diameter at breast height (DBH) in inches were measured. Dry weight of each 5-ft bolt was determined and summed to give stem dry weight above stump. Table 1 gives some simple descriptive statistics on the trees.

**Table 1. Characteristics of the sweetgum data set.**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>Unit of Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. trees</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>No. bolts</td>
<td>824</td>
<td></td>
</tr>
<tr>
<td>Mean DBH</td>
<td>16.2</td>
<td>inches</td>
</tr>
<tr>
<td>DBH range</td>
<td>6.0-32.5</td>
<td>inches</td>
</tr>
<tr>
<td>Mean height</td>
<td>90.5</td>
<td>feet</td>
</tr>
<tr>
<td>Height range</td>
<td>48-128</td>
<td>feet</td>
</tr>
<tr>
<td>Mean stem dry weight</td>
<td>1706</td>
<td>pounds</td>
</tr>
<tr>
<td>Dry weight range</td>
<td>82-6552</td>
<td>pounds</td>
</tr>
</tbody>
</table>

**Equations**

Our first step was to obtain a simple taper model that provides a practical fit of the volume of the bole. We settled on a simple trigonometric transformation that resulted in a three term taper model (Thomas and Parresol 1991) that fit the data as well as multi-termed polynomial models. It has the form:

\[ d^2/D^2 = \beta_1(x-1) + \beta_2(\sin 2\pi x) + \beta_3(\cot \pi x/2) + \epsilon \]  

where \( d \) is diameter inside bark, \( D \) is DBH, and \( \epsilon \) is residual error.

We attempted to construct a specific gravity model for sweetgum, but no significant trend in the data could be found. Consequently, the weight model for sweetgum is the product of the mean specific gravity and the volume integral of the trigonometric taper model. Given \( px,y \) = 29.932, the constant representing weight of a cubic foot, and \( f(x) = n/576 \ D^2/2 \ (x^2 - x_1^2) \) + \( \beta_2(\sin 2\pi x) + \beta_3(\cot \pi x/2) \), where \( n/576 \) is the factor for converting in\(^2\) to ft\(^2\), the following model results upon integrating Equation (1):

\[ w = 29.932 \ \frac{n}{576} \ D^2 H \left[ -\frac{\beta_1(x_u-x_f)}{2} \left( x_1^2 - x_f^2 \right) - \frac{\beta_2}{2\pi} \left( \cos 2\pi x_u - \cos 2\pi x_f \right) + \frac{2\beta_3}{\pi} \left( \ln \sin \pi x_u/2 - \ln \sin \pi x_f/2 \right) \right] + \epsilon \]

Rewriting this expression as a linear model results in:
where \( \ln \) is the natural logarithm, \( P_{21} = \beta_{11}, P_{22} = \beta_{11}/2, P_{23} = -\beta_{12}/(2n), \) and \( P_{24} = 2\beta_{13}/m \). Equations (2) and (3) form a set of linear statistical models with cross-equation constraints. Since no endogenous variables occur on the RHS, estimators such as 2SLS or 3SLS are not required. However, since contemporaneous correlations are hypothesized, this system of two equations meets the SUR requirements and should achieve more efficient estimates over using OLS on Equation (2).

### Restricted SUR Using PROC IML

Equations (2) and (3) can be written in the usual matrix algebra notation as

\[
Y_1 = X_1 \beta_1 + \theta_1
\]

\[
Y_2 = X_2 \beta_2 + \theta_2
\]

where

\[
Y_T = \begin{bmatrix}
(d^2D^2)_1 \\
(d^2D^2)_2 \\
(d^2D^2)_T
\end{bmatrix}
\]

\[
Y_2 = \begin{bmatrix}
(-\frac{576}{29.932nD^4H})_1 \\
(-\frac{576}{29.932nD^4H})_2 \\
(-\frac{576}{29.932nD^4H})_T
\end{bmatrix}
\]

\[
X_T = \begin{bmatrix}
(x-1)_1 (\sin 2nx)_1 (\cos n/2)_1 \\
(x-1)_2 (\sin 2nx)_2 (\cos n/2)_2 \\
(x-1)_T (\sin 2nx)_T (\cos n/2)_T
\end{bmatrix}
\]

The usual assumptions are

\[
E[\theta] = 0 \quad \text{and} \quad E[\theta_i \theta_j'] = \sigma_{ij} \quad i,j = 1,2
\]

Combining all equations into one big model yields

\[
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix} = \begin{bmatrix}
X_1 & 0 \\
0 & X_2
\end{bmatrix}\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} + \begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix}
\]

or alternatively

\[
y = X\beta + \theta
\]

The variances and covariances are unknown and must be estimated, with their estimates being used to form an estimated generalized least squares estimator. To estimate the \( \sigma_{ij} \) we first estimate each equation by least squares residuals \( \hat{\epsilon}_i = Y_i - X_i\beta_i \) and obtain the least squares residuals \( \hat{\epsilon}_i = Y_i - X_i\beta_i \). Consistent estimates of the variances and covariances are then given by

\[
\sigma_{ij} = \frac{1}{(n-k)^{(1/2)}} (r - R\hat{\beta})
\]

where the degrees-of-freedom corrections \( K_k \) and \( K_i \) are the number of coefficients per equation. If we define \( \hat{I} \) as the matrix containing the estimates \( \hat{\theta}_{ij} \) then the unrestricted SUR estimator for \( \beta \) can be written as

\[
\hat{\beta} = (X'X)^{-1}X'y
\]

where \( \Theta \) denotes the Kronecker product. The restricted SUR estimator is obtained by minimizing \( (y - X\hat{\beta})'(2^{-1}\hat{I}X)'(y - X\hat{\beta}) \) subject to the linear restrictions \( R\hat{\beta} = r \). It is given by (Judge et al. 1988; p. 457)

\[
\hat{\beta}^* = \beta + CR'(RCR')^{-1}(r - R\hat{\beta})
\]

where \( C = (X'X)^{-1}I \)

The linear restrictions \( \beta_{21} = \beta_{11}, \beta_{22} = \beta_{11}/2, \beta_{23} = -\beta_{12}/(2n), \) and \( \beta_{24} = 2\beta_{13}/m \), can be written alternatively as \( \beta_{21} + \beta_{22} = 0, \beta_{23} = 0, \beta_{24} = 0, \) and \( 2n\beta_{13} - \beta_{24} = 0 \). Writing these restrictions in the format \( R\hat{\beta} = r \) yields

\[
\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & \frac{1}{2n} & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & \frac{2}{n} & 0 & 0 & 0 & -1
\end{bmatrix}
\]

The covariance matrix of the restricted parameter estimates is calculated as follows

\[
\begin{bmatrix}
\beta_{11} \\
\beta_{12} \\
\beta_{13} \\
\beta_{21} \\
\beta_{22} \\
\beta_{23} \\
\beta_{24}
\end{bmatrix} = \begin{bmatrix}
[0] \\
[0] \\
[0] \\
[0] \\
[0] \\
[0] \\
[0]
\end{bmatrix}
\]
\[ \Sigma_q = C - CR (C R C')^{-1} RC \]

With this background it is relatively easy to program the solution in PROC IML. Appendix A contains the SAS/IML source code. 

Restricted SUR Using PROC SYSLIN

For those users who have access to the SAS/ETS product, the above system of equations can be fitted conveniently and easily using PROC SYSLIN. The cross-equation constraints are imposed using the SRESTRICT statement. Appendix B contains the SAS/ETS program. This program uses data step 'one' from the Program in Appendix A.

ANALYSIS AND RESULTS

If contemporaneous correlation does not exist, least squares applied to the system is fully efficient and there is no need to employ the seemingly unrelated regression estimator. Thus, it is useful to test whether the contemporaneous covariance is 0. In the context of the Density-Integral system, the null and alternative hypotheses for this test are

\[ H_0: \sigma_{12} = 0 \]
\[ H_a: \sigma_{12} \neq 0 \]

An appropriate test statistic is the Lagrange multiplier statistic (Judge et al. 1988; p. 456). In the two equation case this statistic is given by

\[ \lambda = T r^2 \]

where \( T \) is number of observations and \( r^2 \) is the squared correlation

\[ r^2 = \frac{\sigma^2_{ij}}{\delta_{ij} \delta_{ij}} \]

Under \( H_0 \) \( \lambda \) has an asymptotic \( \chi^2 \) distribution with 1 degree of freedom.

Upon applying SUR to the density-integral system of equations the following S-matrix, which contains the \( \sigma^2_q \)'s, was obtained:

\[
\begin{bmatrix}
0.004479 & 0.0001459 \\
0.0001459 & 1.303 \times 10^{-5}
\end{bmatrix}
\]

Using these values to calculate \( \lambda \) gives:

\[ \lambda = 824 \times (0.3647) = 300.5 \]

This large value indicates significant contemporaneous correlation between equations in the system. This suggests SUR should provide gains in efficiency. Table 2 lists the parameter estimates and associated standard errors from OLS and restricted SUR estimation. The standard errors are all smaller under SUR estimation indicating more efficient parameter estimates as was hypothesized.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OLS</th>
<th>SUR</th>
<th>reduction in std err</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b_{11} )</td>
<td>-0.600754 (0.005249)</td>
<td>-0.604647 (0.004413)</td>
<td>16%</td>
</tr>
<tr>
<td>( b_{12} )</td>
<td>0.58825 (0.003618)</td>
<td>0.53831 (0.003043)</td>
<td>16%</td>
</tr>
<tr>
<td>( b_{13} )</td>
<td>0.15087 (0.000275)</td>
<td>0.15505 (0.000218)</td>
<td>21%</td>
</tr>
</tbody>
</table>

CONCLUSIONS

Our investigation of SAS procedures for implementing SUR proved an educational and reassuring exercise. Our understanding of the expression of the matrix forms for restricted and unrestricted SUR improved with the use of PROC IML. The direct application of matrix notation provided us with a confirmation of the applicability of these econometric methods to a forestry problem. It was also reassuring to obtain results that were exactly equivalent to those that were obtained from PROC SYSLIN. In summary, we were pleased to find a unifying approach to the forestry taper-biomass problem directly implementable through SAS procedures.

SAS, SAS/IML, and SAS/ETS are registered trademarks of SAS Institute Inc., Cary, NC, USA.

LITERATURE CITED


AUTHOR CONTACT

Bernard Parresol, Southern Forest Experiment Station Room T-10210 USPS Bldg., 701 Loyola Ave New Orleans, LA 70113 Phone (504) 489-4547
libname source 'directory';

data one; set source.sweetgum;
   y1 = (d/dbh)**2;
   x11 = x - 1;
   pi = 4*atan(1);
   x12 = sin(2*pi*x);
   if x-l then x13=0; else x13 = 1/tan(.5*pi*x);
   y2 = (576*w)/(29.932*pi*dbh*dbh*H);
   x21 = x - x_1;
   x22 = x*x - x_1*x_1;
   x23 = cos(2*pi*x) - cos(2*pi*x_1);
   x24 = log(sin(.5*pi*x)/sin(.5*pi*x_1));

proc iml;
   use one;
   read all;
   X1 = x11 || x12 || x13;   * design matrix for taper eqn;
   X2 = x21 || x22 || x23 || x24;   * design matrix for biomass eqn;
   T = nrow(y1);   * no. of observations;

   start ols;
      K = ncol(X);   * number of variables;
      sinsv = inv(X'X);   * inverse crossproducts;
      bhat = sinsv*X'y;   * ols estimates;
      ehat = y - X*bhat;   * residuals;
      sse = ssq(ehat);   * sum of squared errors;
      sst = ssq(y);   * uncorrected sum of squares total;
      R_sq = 1 - sse/sst;   * R square for regression thru origin;
      sig_sq = sse/(T-K);   * error variance;
      covbhats = sinsv # sinsv;   * covariance of estimates;
      sbhat = sqrt(vecdiag(covbhats));   * standard errors;
      tval = bhat/sbhat;   * t-test for estimates - 0;
      overallf = (sst-sse)/(K*sig_sq);   * F-test for El - ... - Ek - 0;
      p = 1 - probf(overallf,K,T-K);   * F significance probability;
      result1 = sse || sig_sq || overallf || p || R_sq;
      c = (" SSE" "SIGMA_SQ" " F VALUE" " PROB>F" " R_SQ");
      print result1 (icolname=c);
      p = 1 - probf(tval#tval,1,T-K);   * t significance probability;
      result2 = bhat || sbhat || tval || p;
      c = ("PARAMESTIMATE" "STANDARD ERROR" "T FOR HO PARM-O" "PROB>T");
      print result2 (rowname=r colname=c);
   finish;

   print "OLS RESULTS FOR TAPER MODEL";
   y = y1; X = X1; r = ("X11" "X12" "X13"); run ols;
   e1 = ehat;   * ols residuals for taper eqn;
   print / "OLS RESULTS FOR BIOMASS MODEL";
   y = y2; X = X2; r = ("X21" "X22" "X23" "X24"); run ols;
   e2 = ehat;   * ols residuals for biomass eqn;
   K1 = ncol(X1); K2 = ncol(X2);
   e1 = e1/sqrt(T-K1);
   e2 = e2/sqrt(T-K2);
ihat~ e1 || e2;
S = ehat*ehat;
print / "RESTRICTED SEEMINGLY UNRELATED REGRESSIONS RESULTS";
print "MATRIX OF CONTEMPORANEOUS COVARIANCES, i.e. SIGMA MATRIX";
print S;
y = y1 // y2;
X = block(X1,X2);
free y1 y2 X1 X2 e1 e2 ehat;
PHI = inv(S) @ I(T);
C = inv(X'*PHI*X);
bhat= C*X'*PHI*y;
free PHI;
RD = (1 0 0 1 0 0 0,
      .5 0 0 0 -1 0 0,
      0 .159154943 0 0 1 0,
      0 0 .636619772 0 0 0 -1);
T = (0,0,0,0);
bstar= bhat + C*RD*inv(RD*C*RD')*(r - RD*bhat);
covar = C - C*RD*inv(RD*C*RD')*RD*C;
sdbstar = sqrt(vecdiag(covar));
tval = bstar/sdbstar;
p = (1-probt(abs(tval),1))*2;
results = bstar || sdbstar || tval || p;
r = ("X11" "X12" "X13" "X21" "X22" "X23" "X24");
c = ("PARAMETRESTIMATE" "STANDARD ERROR" "T FOR HO PARM=0" "PROB>|T|"");
print "RESTRICTED SUR ESTIMATES";
print results (irownname=r colname=c);
quit;
run;

APPENDIX B

proc syslin data=one sur;
taper: model y1 = x1 x2 x3/noint;
biomass: model y2 = x21 x22 x23 x24/noint;
srestrict taper.x11 + biomass.x21 = 0,
     .5*taper.x11 - biomass.x22 = 0,
     .159154943*taper.x12 + biomass.x23 = 0,
     .636619772*taper.x13 - biomass.x24 = 0;
run;