Forest Science, Vol. 35, No. 1, pp. 105–125.

Model Fitting Under Patterned Heterogeneity of Variance

TIMOTHY G. GREGOIRE
MICHAEL E. DYER

ABSTRACT. Two approaches toward fitting regression models with multiplicative error heteroscedasticity recur in the forestry, ecology, and statistical literature. One includes the estimation of the heterogeneity in the fitting process. The alternative approach entails the use of variance estimators that are robust to the error variance heterogeneity. Under suitable conditions, the former method offers nonnegligible gains in efficiency, whereas the robust alternatives provide accurate assessment of ordinary least squares estimators even in the presence of heteroscedasticity. The performance of both approaches are examined and contrasted, and suggestions for future applications and research are made on the basis of these results. FOR. SCI. 35(1):105–125.

ADDITIONAL KEY WORDS: Weighted least squares, heteroscedasticity, jackknife, bootstrap.

WE CONSIDER FITTING A LINEAR REGRESSION MODEL to observations that are uncorrelated but have nonuniform variance. The pattern of heterogeneity is assumed known, but its degree or level is not. The relevance of this model to forestry was established by Gedney and Johnson (1959) and Cunia (1964), and it has been used subsequently in many volume and biomass modeling efforts. As is well known, the application of ordinary least squares (OLS) methods in the presence of variance heterogeneity[1] results (1) in coefficient estimators that do not have minimum variance in the class of linear, unbiased estimators, and (2) in biased estimators of the variances of the estimated coefficients. Consequently, the coverage probabilities for interval estimates of model coefficients are perturbed away from their nominal values, as are the actual significance levels of hypothesis tests relating to the model parameters. For the model developer, the customary tests used in stepwise fitting techniques or other variable selection methods that aim to provide a parsimonious model, are not conducted with the risks of Type I and Type II error that one presumes, nor are the actual error rates readily discernible. For the model user, predictions come from a more disperse sampling distribution than one is led to believe. Cunia (1965) discussed the importance of including the error of volume equations, along with sampling error, when assessing the reliability of forest inventory estimates. The component of error due to the volume equation is difficult to discern if the vari-

Timothy G. Gregoire is Assistant Professor of Forest Biometrics, Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 and at the time this paper was written, Michael E. Dyer was Graduate Research Assistant, Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061. Dyer is presently employed as a research scientist with Westvaco Corporation, Summerville, SC 29484. The authors are grateful to Harold E. Burkhart and industrial forestry cooperators at Virginia Polytechnic Institute and State University for providing the loblolly pine stem measurements used in this investigation. The thorough and helpful reviews provided by T. Cunia and Donald Seegrist are also gratefully acknowledged. Manuscript received April 1, 1988.

[1] We presume that the other, usual assumptions about the behavior of the model errors are satisfied.

ance estimators of the volume equation have a bias of unknown magnitude. Variance heterogeneity has also been faulted for leading to coefficient estimates with signs that were counterintuitive or antithetical (Box and Hill 1974).

Two approaches toward fitting heterogeneous error models recur in the forestry, ecology, and statistical literature. In the first, the heterogeneity itself is modeled, and its effect is included in the estimation of the model parameters (cf. Cunia 1964, Chew 1970, Schreuder and Swank 1973, Fuller and Rao 1978, McClure et al. 1983, Carroll and Ruppert 1985). In the second, variance heterogeneity is ignored, but the covariance matrix of the fitted coefficients is estimated by techniques, other than OLS, that are resistant to error variance heterogeneity (cf. White 1980, Efron and Gong 1983, Hinkley 1977, Wu 1986).

We examine the performance of alternative procedures that can be used with both approaches in the case of a simple linear model. The first part of the study consists of an empirical comparison of methods to model heterogeneity of variance as applied to two sets of tree volume data. The second part of the study uses simulation to analyze the behavior of a number of alternative variance estimators that are robust to departures from uniformity of error variance.

MODEL AND NOTATION

The model under consideration is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$

where i indexes observation, $i = 1, \ldots, n$, and where $E[\epsilon_i|X_i] = 0$, $V[\epsilon_i|X_i] = E[\epsilon_i^2|X_i] = \sigma_i^2 = \sigma^2 X_i^\omega$. Thus the conditional variance of Y is proportional to X raised to a power, ω. Furthermore, observations are uncorrelated: $E[(\epsilon_i|X_i)(\epsilon_j|X_j)] = 0$, for distinct observations i,j. All but $\{Y_i, X_i\}$ are unknown and unobservable. Interest is focused on $\underline{\beta}^*$ and Var$[\underline{\beta}^*]$, for some estimator $\underline{\beta}^*$ of $\underline{\beta} = [\beta_0, \beta_1]'$, and on an estimator of Var$[\underline{\beta}^*]$ denoted by var$[\underline{\beta}^*]$. We let $\underline{Y} = [Y_1, \ldots, Y_n]'$, $\underline{Z}_i' = [1\ X_i]$, $Z = [\underline{Z}_1, \ldots, \underline{Z}_n]'$, $\underline{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]'$, and $\Omega = E[\underline{\epsilon}\underline{\epsilon}'] = \sigma^2 V$, where $V = \text{diag}\{X_i^\omega\}$. The aggregate model is expressed $\underline{Y} = Z\underline{\beta} + \underline{\epsilon}$.

Hartley and Jayatillake (1973) and others[2] treat linear model estimation under a less specific error variance heterogeneity. The pattern of multiplicative heterogeneity embodied in (1), namely $\sigma_i^2 = \sigma^2 X_i^\omega$, is well-established in volume, weight, and biomass relationships common to forestry and ecological modeling. This more specific error structure is a source of knowledge that, presumably, can be used to advantage over a completely unspecified heterogeneity, namely $\sigma_i^2 \neq \sigma_j^2$. Box and Hill (1974) and Seber (1977) have treated the case where $\sigma_i^2 = g(E[Y_i])$.

CONSEQUENCES OF HETEROSCEDASTICITY

The optimal unbiased estimator of $\underline{\beta}$ under mean square error loss is

$$\underline{\beta}_G = (Z'V^{-1}Z)^{-1}Z'V^{-1}\underline{Y} \tag{2}$$

For any arbitrary but *fixed* specification of the weighting matrix, say V^*,

[2] Jacques, Mather, and Crawford (1968), Horn, Horn, and Duncan (1975), Fuller and Rao (1978), Cragg (1983), Deaton, Reynolds, and Myers (1983).

$$\underline{\beta}_G^* = (Z'V^{*-1}Z)^{-1}Z'V^{*-1}\underline{Y} \tag{3}$$

also is unbiased but its covariance

$$\begin{aligned}\mathrm{Var}(\underline{\beta}_G^*) &= (Z'V^{*-1}Z)^{-1}(Z'V^{*-1}\Omega V^{*-1}Z)(Z'V^{*-1}Z)^{-1} \\ &= \sigma^2(Z'V^{*-1}Z)^{-1}(Z'V^{*-1}VV^{*-1}Z)(Z'V^{*-1}Z)^{-1}\end{aligned} \tag{4}$$

exceeds

$$\begin{aligned}\mathrm{Var}(\underline{\beta}_G) &= (Z'\Omega^{-1}Z)^{-1} \\ &= \sigma^2(Z'V^{-1}Z)^{-1}\end{aligned} \tag{5}$$

by a positive definite matrix. This results from Aitken's Theorem (cf. Theil 1971). Bloomfield and Watson (1975) derived the bounds on the variance of an arbitrary linear combination, $L'\underline{\beta}_G^*$, relative to the variance of $L'\underline{\beta}_G$. Bloch and Moses (1988) generalized their results slightly for the simple linear regression model only.

When V^* is the identity matrix, I_n, $\underline{\beta}_G^*$ is the OLS estimator, which, for convenience, we denote separately as

$$\tilde{\underline{\beta}} = (Z'Z)^{-1}Z'\underline{Y} \tag{6}$$

The OLS estimator has dispersion

$$\mathrm{Var}(\tilde{\underline{\beta}}) = (Z'Z)^{-1}(Z'\Omega Z)(Z'Z)^{-1} \tag{7}$$

Note that

$$\mathrm{Var}' = \sigma^2(Z'Z)^{-1} \tag{8}$$

is a biased approximation to $\mathrm{Var}(\tilde{\underline{\beta}})$. Moreover the bias is negative whenever $\omega > 0$, $X_i > 1$, $i = 1, \ldots, n$, in the sense that

$$\mathrm{Var}' - \mathrm{Var}(\tilde{\underline{\beta}}) = \sigma^2(Z'Z)^{-1}(Z'[I_n - V]Z)(Z'Z)^{-1}$$

is a negative definite matrix. In this case, all elements of Var′ are smaller than corresponding elements of $\mathrm{Var}(\tilde{\underline{\beta}})$, although this may not obtain when the usual estimator of Var′ is used. This fact is apparent in the simulation results presented later.

Because $\mathrm{Var}(\tilde{\underline{\beta}})$ exceeds $\mathrm{Var}(\underline{\beta}_G)$, then for an arbitrary linear combination $L'\underline{\beta}$, $\mathrm{Var}(L'\tilde{\underline{\beta}})$ exceeds $\mathrm{Var}(L'\underline{\beta}_G)$. Hence the use of OLS estimators to derive predictions deleteriously affects the reliability of predictions.

We note, too, that the usual OLS estimator of σ^2,

$$\begin{aligned}\tilde{\sigma}^2 &= (\underline{Y} - Z\tilde{\underline{\beta}})'(\underline{Y} - Z\tilde{\underline{\beta}})/(n-2) \\ &= \underline{\epsilon}'(I_n - Z(Z'Z)^{-1}Z')\underline{\epsilon}/(n-2)\end{aligned} \tag{9}$$

is biased by an amount equal to

$$\begin{aligned}E[\tilde{\sigma}^2] - \sigma^2 &= \sigma^2[\mathrm{trace}\{(I_n - Z(Z'Z)^{-1}Z')V\}/(n-2) - 1] \\ &= \frac{\sigma^2}{n}\left[\frac{n-1}{n-2}\sum_{i=1}^{n} X_i^{\omega} - \frac{n}{(n-2)S_x^2}\sum_{i=1}^{n} x_i^2 X_i^{\omega} - n\right]\end{aligned}$$

where $x_i = X_i - \overline{X}$ and S_x^2 is the mean corrected sum of squares. (The estimator of σ^2 based on $\underline{\beta}_G^*$ is similar, and it is given in the Appendix.) The usual OLS estimator of $\mathrm{Var}(L'\tilde{\underline{\beta}})$ is $\tilde{\sigma}^2L'(Z'Z)^{-1}L$, which is biased due to both the bias of $\tilde{\sigma}^2$ in estimating σ^2 and that of Var′ in approximating $\mathrm{Var}(\tilde{\underline{\beta}})$. Swindel (1968) has derived the attainable bounds on the bias of this variance estimator. Since these bounds depend on V, they are not feasible in the sense of being calculable for a specific problem. Sathe and Vinod (1974) provide an illuminating approximation to these bounds.

Consider now the case for an *estimated,* rather than fixed, specification of the weighting matrix. If this estimated matrix is denoted by $\tilde{V}$, then

$$\underline{\tilde{\beta}}_G = (Z'\tilde{V}^{-1}Z)^{-1}Z'\tilde{V}^{-1}\underline{Y} \tag{10}$$

is a feasible GLS estimator (FGLS). Unless stipulated differently, we assume that $\tilde{V}$ is a consistent estimator of V. For certain estimators of V, $\underline{\tilde{\beta}}_G$ is asymptotically equivalent to $\underline{\beta}_G$, and Carroll and Ruppert (1982) showed that this equivalency extends to a broad class of robust estimators of V. Kakwani (1967) proved that $\underline{\tilde{\beta}}_G$ remains unbiased for $\underline{\beta}$ provided that the errors follow a symmetric distribution and other, mild regularity conditions are satisfied. Don and Magnus (1980) later extended this result to iterative estimators of the same form as the two-step estimator shown in (10). Under error normality, $\mathrm{Var}(\underline{\tilde{\beta}}_G)$ is related directly to the precision with which ω is estimated (Rothenberg 1984), and Davidian and Carroll's (1987) results indicate that the precision of an estimator of ω is relatively unaffected by the magnitude of σ.

Under normality, the joint maximum likelihood estimator (ML) of $\underline{\beta}$ and Ω is asymptotically efficient, also. With small samples, the relative performances of feasible GLS estimators and the MLE are unclear, because the uncertainty in the estimate of V contributes to the variability of the estimator of $\underline{\beta}$. Despite this added variability, Carroll and Ruppert (1986) recommend modeling the heterogeneity in preference to using OLS with a robust estimate of the dispersion matrix. Taylor's (1978) results offer cogent support of this view, although Taylor considered a slightly different pattern of heteroscedasticity. Others are less sanguine about this approach.

MODELING HETEROSCEDASTICITY

A number of feasible GLS estimators are possible, differing with respect to the estimator of ω (in $\tilde{V}$) used in the first of the two-step estimation processes. These estimators can be classified as based on grouped, or perhaps replicated, X values or as based on residuals from a preliminary fit of the model. Alternatives to feasible GLS estimation are procedures that estimate ω jointly with the other unknowns, e.g., maximum likelihood estimation.

Given the multiplicative heteroscedasticity, as specified in Ω, it is widely known that $\underline{\beta}_G$ can be obtained by OLS applied to the transformed observations $\{Y_i^*, \underline{Z}_i^{*\prime}\}$, where $Y_i^* = Y_i X_i^{-\omega/2}$ and $\underline{Z}_i^{*\prime} = [X_i^{-\omega/2}, X_i^{1-\omega/2}]$. Likewise, the feasible GLS estimator is equivalent to OLS applied to the transformed data using an estimate of ω. It does not seem possible, however, that any transformation of model (1) that does not involve X_i can effect an optimal solution. For example, the family of power transformations of the dependent variable put forth by Box and Cox (1964) and Box and Hill (1974) fall into this category. The latter transformation method was designed to find an optimal power transformation of Y_i in the case where $\sigma_i^2 \propto E[Y_i^\tau]$. The success of this transformation in the case considered here would depend on how well $[\underline{Z}_i'\underline{\beta}]^\tau$ approximates X_i^ω. Furthermore, any transformation of just the predictor variables, such as the type proposed by Box and Tidwell (1962), cannot correct heterogeneity of error variance of the type we consider.

In many statistics textbooks and journal articles it is suggested that weights be estimated from replicated data at each X value; see, for example, Draper and Smith (1981), Myers (1986), Deaton et al. (1983). Forestry data rarely are collected by a design or plan that produces replicated data points, a circumstance which limits the utility of this approach in many forestry modeling efforts. Moreover it is not clear that replication is so important when trying to model patterned heterogeneity, because variance is continuously related to X.

Schumacher and Chapman (1942) present one of the earliest applications of weighted least squares in a forestry context. Gedney and Johnson (1959) apparently were the first to use weighted least squares to improve the precision of a tree volume model in which volumes become increasingly disperse as tree size increases. A combination of regression and graphical techniques was used to estimate the appropriate weights. Schreuder and Swank (1973) estimated the weight parameter, ω, in a manner similar to Cunia (1964), by sorting the observations on X, dividing the observations into intervals of equal width, and regressing the group sample variances on the average value of X in the interval. In other applications, the number of observations in a group has been held constant, thereby forming intervals of unequal width.

There is an arbitrariness about this method of estimating ω that derives from the manner in which the range of X is subdivided (cf. McClure and Czaplewski 1987). If the variance of Y increases with X, then the variance within an interval increases directly with the interval width. This suggests that the intervals should be uniform and reasonably short with respect to the range of X. But as the interval width decreases, so too does the number of observations available to estimate the variance of Y within the interval. Since small sample sizes can yield very imprecise estimates of variance, researchers have typically chosen some minimum number of observations within an interval in order to include that interval in the estimation of the variance function.

When sample data are skewed with respect to the X variable, as they are apt to be if collected by sample survey rather than design, the relationship between interval width and number of observations within an interval will not remain constant. Consequently, if interval widths are made equal, the number of observations in each interval will vary. In particular, it is well established that the frequency distribution of tree volume becomes positively skewed with time (Ford 1975). The implication for tree volume models is that intervals over the larger values of X may contain few observations. If a minimum sample size is established for fixed interval width, then many of the data corresponding to larger trees may not be utilizable. One may combine intervals to achieve a minimum frequency requirement, but the advantage of equal size intervals is forfeited by this strategy, and the comparability of variance estimates based on differing numbers of observations is suspect.

The effect of sample size on variance estimates and the consequences of skewed data is evident in McClure et al. (1983). In this study, trees were sorted and grouped into intervals of width 1000 (in.^{2}*ft). With the loblolly pine data, the number of classes decreased from 89, when minimum within-class sample size was fixed at 15, to 11, when minimum within-class sample size was fixed at 75. The estimates of ω, displayed in their Table 1 for different minimum sample sizes, ranged from 1.61 to 1.13 for the loblolly pine data. By comparing their Figure 3 to Figure 1, it is apparent that the decline in the number of classes was most severe in the large tree classes, yet it is precisely those size classes that are apt to be most revealing of the trend and the magnitude of variance heterogeneity. The loss of information in the large size classes seems unavoidable with this manner of modeling heterogeneity.

Since the OLS residuals derive their properties from the corresponding errors in the model, it seems reasonable to use them directly to estimate ω. Let e_i denote the OLS residual corresponding to the ith observation and let h_{ij} denote the ijth element of the hat matrix $H = Z(Z'Z)^{-1}Z'$ (Belsley, Kuh, and Welsch 1980). Then,

$$\begin{aligned} e_i &= Y_i - \underline{Z}_i'\underline{\hat{\beta}} \\ &= \epsilon_i - \sum_{j=1}^{n} h_{ij}\epsilon_j \end{aligned}$$

and, for the homoscedastic model, $\mathrm{Var}(e_i) = E[e_i^2] = (1 - h_{ii})\sigma^2$. This suggests that the approximate relation

$$\begin{aligned} E[e_i^{*2}] &\cong \sigma_i^2 \\ &\cong \sigma^2 X_i^{\omega} \end{aligned} \tag{11}$$

where $e_i^* = e_i^2/(1 - h_{ii})$ may serve as a suitable model for the heterogeneous error variance. In particular, linearize (11) by a logarithmic transformation and consider the model

$$R_i^* = \underline{S}_i'\underline{\alpha} + \delta_i \tag{12}$$

where $R_i^* = \ln(e_i^{*2})$, $\underline{\alpha} = [\alpha_0, \alpha_1]' = [\ln(\sigma^2), \omega]'$, $\underline{S}_i' = [1, \ln(X_i)]$, and δ_i is error. An estimator of ω is the slope coefficient estimator, $\tilde{\omega} = \tilde{\alpha}_1$, in an OLS fit of (12). A feasible GLS fit of (1) entails fitting (12) by OLS to generate an estimate of the weight parameter, $\tilde{\omega}$, and then fitting (1) by weighted least squares using the reciprocal $X_i^{\tilde{\omega}}$ as weights.

Absolute and squared OLS residuals have been used in numerous other schemes to estimate heterogenous variances. For example, minimum norm quadratic unbiased estimators (Rao 1970) are $\sigma^2_{i,\mathrm{MINQUE}} = \bar{h}_{i1}^2 e_1^2 + \ldots + \bar{h}_{in}^2 e_n^2$, where $\bar{h}_{ij}$ is the ijth element of the projection matrix $\bar{H} = (I_n - Z(Z'Z)^{-1}Z')^{-1}_{-}$. AUE (for "almost unbiased estimators") can be derived as $\sigma^2_{i,\mathrm{AUE}} = e_i^2/\bar{h}_{ii}$ (Horn, Horn, and Duncan 1975). Squared residuals are integral to the estimators developed, too, by Hildreth and Houck (1968), Chew (1970), Goldfield and Quandt (1972), Amemiya (1977), and Jobson and Fuller (1980). When σ_i^2 does not depend on $\underline{\beta}$, many FGLS estimators of $\underline{\beta}$ which use weights based on untransformed, squared residuals are asymptotically equivalent to maximum likelihood estimation (Davidian and Carroll 1987).

Harvey (1976) suggested that $\sigma_i^2 = \sigma^2 X_i^{\omega}$ be re-expressed as $\sigma_i^2 = \exp(\underline{S}_i'\underline{\lambda})$, where $\underline{\lambda} = [\ln(\sigma^2), \omega]'$, and that the transformed, squared OLS residuals be modeled as

$$R_i = \underline{S}_i'\underline{\lambda} + u_i \tag{13}$$

where $R_i = \ln(e_i^2)$, and $u_i = \ln(e_i^2/\sigma_i^2)$. He showed that the OLS estimator $\underline{\tilde{\lambda}} = [\tilde{\lambda}_1, \tilde{\lambda}_2]' = (S'S)^{-1}S'\underline{R}$, is a consistent estimator of $\underline{\lambda}$ if 1.2704 is added to the first element.

The log-likelihood function for model (1) under error normality is

$$\begin{aligned}\ln L &= \gamma + \frac{1}{2}\ln|\Omega^{-1}| - \frac{1}{2}\underline{\epsilon}'\Omega^{-1}\underline{\epsilon} \\ &= \gamma - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\ln(X_i^{\omega}) \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2/(\sigma^2 X_i^{\omega}) \end{aligned} \tag{14}$$

where $\gamma = -(n/2)\ln(2\pi)$. Its derivatives with respect to the four unknown parameters, when evaluated at zero, provide the maximum likelihood (ML) estimating equations:

$$\partial\ln L/\partial\beta_0 = 0 = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)/(\sigma^2 X_i^{\omega})$$

$$\partial\ln L/\partial\beta_1 = 0 = \sum_{i=1}^{n}X_i(Y_i - \beta_0 - \beta_1 X_i)/(\sigma^2 X_i^{\omega})$$

$$\partial\ln L/\partial\sigma^2 = 0 = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2/(\sigma^4 X_i^{\omega})$$

$$\partial\ln L/\partial\omega = 0 = -\frac{1}{2}\sum_{i=1}^{n}\ln(X_i^{\omega}) + \frac{1}{2}\sum_{i=1}^{n}\ln(X_i)(Y_i - \beta_0 - \beta_1 X_i)^2/X_i^{\omega}$$

Magnus (1978) demonstrated that iterative solution of the estimating equations converge to the maximum likelihood solution, $\underline{\beta}_{ML}$. Often a grid search conditioned on values of ω will find the convergent solution quickly. A consistent estimator of the asymptotic Var($\underline{\beta}_{ML}$) is given by the inverse of the information matrix evaluated at the estimated parameter values (Kmenta 1986). Amemiya (1977) showed that the ML estimator of ω is asymptotically fully efficient. Its asymptotic variance, which can be computed, is

$$\mathrm{Var}(\omega_{ML}) = 2\left\{\sum_{i=1}^{n}[\ln(X_i) - \overline{\ln(X_i)}]^2\right\}^{-1} \tag{15}$$

where $\overline{\ln(X_i)}$ is the mean $\ln(X_i)$ value. Harvey (1976) developed the likelihood ratio test statistic for the homogeneity hypothesis, namely H: $\omega = 0$. Meng and Tsai (1986) applied ML to model (1) in the case where X is tree dbh.

Another alternative method to obtain estimates of the model coefficients is minimization of the objective function Δ:

$$\Delta = (\underline{Y} - Z\underline{\beta})'V^{-1}(\underline{Y} - Z\underline{\beta})/\underline{Y}'V^{-1}\underline{Y} \tag{16}$$

The statistical properties of the resulting estimators are unclear, however, which limits the utility of this method.

EMPIRICAL COMPARISON OF METHODS TO MODEL ERROR HETEROSCEDASTICITY

In order to glean insight into the relative merits and disadvantages of the procedures just described, a combined-variable volume equation (Spurr 1952) was fitted to two independent data sets. The model is as specified in (1) with Y_i = total stem, outside-bark cubic volume and X_i = squared dbh times total tree height. The model was fitted by (i) OLS; (ii) GLS with ω

preset to 1.5, denoted as GLS1.5; (iii) feasible GLS (FGLS1) with ω estimated by (12); (iv) feasible GLS (FGLS2) with ω estimated by Harvey's estimator, (13); (v) by maximum likelihood (ML)[3].

The first data set consisted of measurements from 209 loblolly pine (*Pinus taeda* L.) trees from natural stands in the Piedmont region of Virginia and the Coastal Plain region of Virginia and North Carolina (USA). The outside-bark volume (ft^3) of each tree was determined by sectioning at 4 ft (1.22 m) intervals and applying Smalian's rule to each section. Further details of the data collection effort are described in Burkhart et al. (1972). Dbh values ranged from 4.6 in. to 14.1 in. (11.7 – 35.8 cm) and total heights ranged from 32.4 ft to 88.5 ft (9.9 – 27.0 m). The increasing dispersion of stem volume is evident in Figure 1.

The second data set consisted of measurements of 91 planted, red pine (*Pinus resinosa*) trees in Connecticut (USA). The standing trees were measured with a diameter tape at 1 ft (0.30 m) intervals to a height of 6 ft (1.8 m), and thereafter at 4 ft (1.22 m) intervals extending into the base of the crown. Tree volume (ft^3), outside-bark, was computed with an interpolating cubic spline function fitted to the measured stem values, as described in Gregoire et al. (1986). Red pine dbh values ranged from 8.4 in. to 18.2 in. (21.3 – 46.2 cm) and total heights ranged from 59.9 ft to 87.4 ft (18.3 – 26.6 m). The red pine stem volumes displayed in Figure 2 show a pattern of increasing variance that is similar to, albeit more pronounced, than that of the loblolly volumes.

The results from the five procedures are summarized in Table 1 for both species. For loblolly pine, the coefficient estimates are practically indistin-

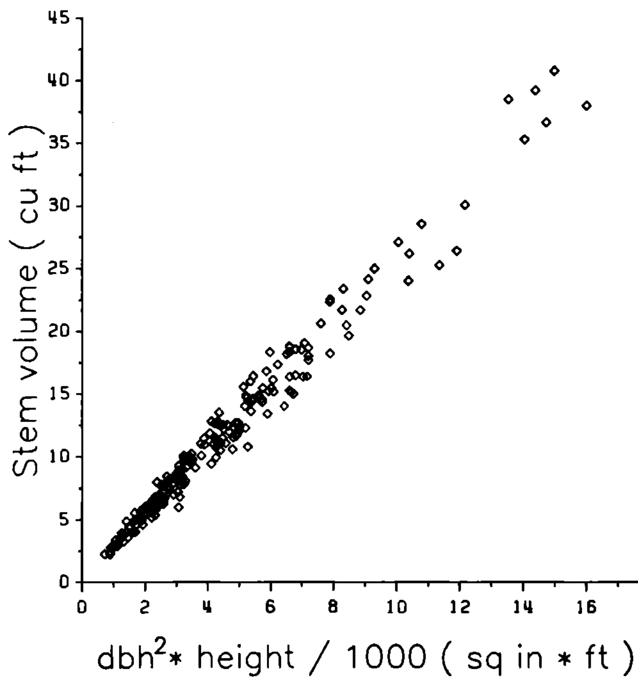

FIGURE 1. Natural stand loblolly pine stem volumes.

[3] An interactive FORTRAN program to compute the ML estimates is available from the authors on request.

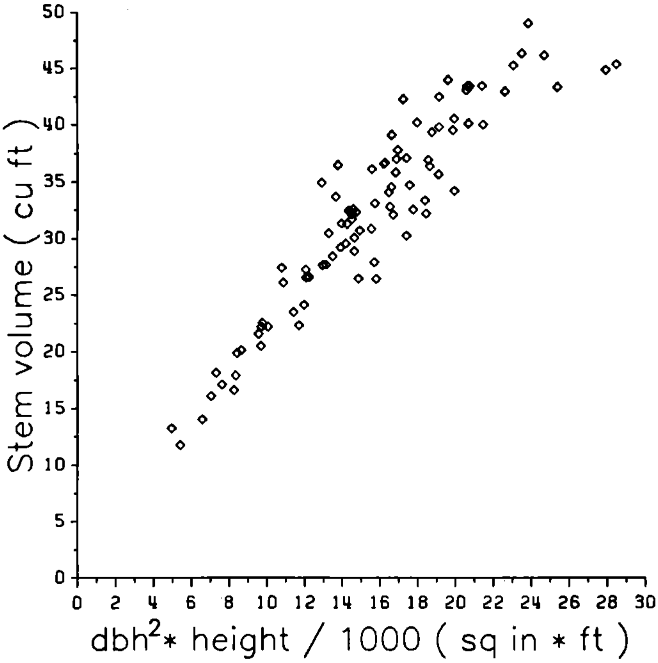

FIGURE 2. Planted stand red pine stem volumes.

guishable among methods ii–v, whereas somewhat greater variation is evident among the coefficient estimates derived for the red pine volume equation. For both species, the OLS estimates of β_0 are elevated above the estimates from the competing methods, whereas the estimates of β_1 are slightly depressed. There should not be a systematic departure of OLS coefficient estimates from GLS and FGLS estimates in general.

The OLS standard errors in Table 1 are the corresponding roots of var($\tilde{\beta}$) $= \tilde{\sigma}^2(Z'Z)^{-1}$ with $\tilde{\sigma}^2$ as defined in (9). The GLS1.5 standard errors are derived from $\sigma^{2*}(Z'V^{*-1}Z)^{-1}$, where $V^* = \text{diag}\{X_i^{1.5}\}$ and σ^{2*} is as defined in the Appendix. The FGLS standard errors are derived in an analogous fashion but with the corresponding estimates of ω in place of the fixed value 1.5 (see Appendix). The ML standard error estimates are computed by (15) and by the equations shown also in the Appendix.

The variance estimators by which the standard errors in Table 1 are computed are all biased in finite samples; however, all but the OLS estimators are consistent. The bias in the OLS estimator derives jointly from the bias in $\tilde{\sigma}^2$ as an estimator of σ^2 and from the approximation to Var($\tilde{\beta}$) by $\sigma^2(Z'Z)^{-1}$. When $X_i > 1$ and $\omega > 0$, the bias in var($\tilde{\beta}_0$) and in var($\tilde{\beta}_1$) is negative.

For both species, the OLS estimate of se(β_0) is substantially larger than the estimates from other methods. For the loblolly pine data, the range in se(β_0) values from methods ii–v is small in relation to their range from the OLS value. This is not the case for the estimates of se(β_0) based on the red pine data. Approximately unbiased variance estimates of Var ($\tilde{\beta}_0$) can be derived by Wu's weighted jackknife estimator (Wu 1986), which is discussed more fully in the next section. For the loblolly pine data, the approximately unbiased standard error estimate is 0.16259 ft^3 and for the red pine data, is 1.24696 ft^3. For both species, the downward bias is slightly more than 13%.

In contrast to the above pattern, the OLS estimates of se(β_1) are much

TABLE 1. Parameter estimates, estimated standard errors, and fit statistics for combined variable volume equation. Parenthesized values indicate that the parameter value was predetermined, not estimated.

					Loblolly pine data (n = 209)					
Procedure		β_0	se (β_0)	β_1	se (β_1)	ω	se (ω)	σ	|error|[a]	Observed log-likelihood
(i)	OLS	0.658	0.14110	0.00249	0.000026519	(0.00)	(0.00)	1.13285	7.1	−323.63
(ii)	GLS1.5	0.456	0.08103	0.00254	0.000027717	(1.50)	(0.00)	0.00186	12.7	−266.72
(iii)	FGLS1	0.429	0.07555	0.00255	0.000028090	1.70	0.24147	0.00080	13.7	−265.03
(iv)	FGLS2	0.430	0.07572	0.00255	0.000028077	1.70	0.24143	0.00082	13.7	−265.06
(v)	ML	0.413	0.07190	0.00256	0.000028229	1.84	0.15062	0.00046	14.4	−263.69
					Red pine data (n = 91)					
Procedure		β_0	se (β_0)	β_1	se (β_1)	ω	se (ω)	σ	|error|[a]	Observed log-likelihood
(i)	OLS	6.74	1.07753	0.00163	0.000066038	(0.00)	(0.00)	3.07087	7.4	−232.24
(ii)	GLS1.5	4.26	0.76462	0.00179	0.000055812	(1.50)	(0.00)	0.00215	8.4	−225.06
(iii)	FGLS1	4.93	0.86518	0.00175	0.000059183	1.01	0.61116	0.02266	7.9	−226.46
(iv)	FGLS2	4.92	0.86407	0.00175	0.000059146	1.02	0.61027	0.02209	7.9	−226.44
(v)	ML	3.65	0.64911	0.00184	0.000051710	2.07	0.42359	0.00013	9.3	−223.39

[a] Average absolute error as a percentage of arithmetic mean cubic volume in the case of OLS and as a percentage of the weighted mean cubic volume in the other cases.

closer in value to the estimates provided by competing methods. The OLS estimate is slightly less than the others for loblolly pine, whereas that for red pine is larger. These seemingly dissonant results are explicable: the variance of the OLS estimator of β_1 exceeds the asymptotic variances of the alternative estimators, yet the OLS variance estimator is negatively biased. Therefore, the negative bias may yield estimates that, depending on the sample, are greater or lesser than those provided by the asymptotically correct variance estimators. Using the Wu weighted jackknife to provide an approximately unbiased basis for comparison, the bias in the OLS estimate of se(β_1) was -41% and -22% for the loblolly and red pine data, respectively.

The two FGLS methods provided essentially equivalent estimates. Apparently the correction for influence embodied in FGLS1 has little effect. The FGLS estimates of ω for both species were lower than the corresponding ML estimate. For the red pine data, the ML estimate of ω is twice that of the FGLS estimates. We conjecture that this may reflect the imprecision that is indicated for the FGLS estimates of ω for these data. Harvey (1976) demonstrates that the FGLS2 estimate of se(ω) should be roughly 60% larger than the corresponding ML estimate. That phenomenon is evident in the results for both species.

As expected, the OLS estimates of σ are substantially larger than the estimates resulting from other methods. Because estimates of $\underline{\beta}$ do not change very much among methods ii–v, estimates of σ decrease directly with increasing estimates of ω.

For both species, the average absolute residual, expressed as a percentage of the mean tree volume, is least for OLS and is greatest for ML. Conversely, the log likelihood, evaluated at the estimated parameter values, was least for OLS. The FGLS and GLS1.5 estimators had observed log-likelihoods that were just slightly less than that for the ML fit.

ROBUST ESTIMATORS OF Var($\tilde{\underline{\beta}}$)

None of the robust estimates of Var($\tilde{\underline{\beta}}$) outlined in this section rely on a specific pattern of variance heterogeneity. All were either derived from a class of resampling estimators (cf. Efron 1982, Efron and Gong 1983) or can be directly related to estimators in this class.[4]

White (1980) noted that, given conditions regulating the asymptotic behavior of Z′Z, the OLS residuals can be used to derive

$$\begin{aligned} \mathrm{var}_W(\tilde{\underline{\beta}}) &= (Z'Z)^{-1}(Z'\mathrm{diag}\{e_i^2\}Z)(Z'Z)^{-1} \\ &= (Z'Z)^{-1}\left(\sum_{i=1}^{n} e_i^2 \underline{Z}_i\underline{Z}_i'\right)(Z'Z)^{-1} \qquad (17) \end{aligned}$$

as a consistent estimator of Var($\tilde{\underline{\beta}}$). As before, the errors, ϵ_i, are assumed to be mutually uncorrelated, and e_i is the OLS residual. Notice that (17) is an empirical counterpart to $(Z'Z)^{-1}(Z'\mathrm{diag}\{E[\epsilon_i^2]\}Z)(Z'Z)^{-1}$ which is a re-expression of (7), the OLS covariance matrix.

Jackknifing methods to estimate Var($\tilde{\underline{\beta}}$) have been discussed by Hinkley (1977) and more recently by Wu (1986). In our discussion, we restrict attention to "delete-one" jackknifing plans, only, because "delete-one" applications are more prevalent and because "delete-many" plans become cumbersome.

---

[4] See Efron (1982) and Gregoire (1984) for an introduction to these resampling plans and the accompanying terminology.

The ordinary jackknife estimator of Var($\underline{\tilde{\beta}}$)

$$\mathrm{var}_J(\underline{\tilde{\beta}}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\underline{\beta}_i - \underline{\beta}_J)(\underline{\beta}_i - \underline{\beta}_J)', \tag{18}$$

where $\underline{\beta}_i$ is the ith pseudovalue and $\underline{\beta}_J$ is the jackknife estimator of $\underline{\beta}$, which is calculated as the arithmetic average pseudovalue. It can be shown that $\mathrm{var}_J(\underline{\tilde{\beta}})$ can be re-expressed in a manner similar to (17):

$$\mathrm{var}_J(\underline{\tilde{\beta}}) = \frac{n-1}{n} (Z'Z)^{-1}(Z'[\mathrm{diag}\{q_i\} - Q]Z)(Z'Z)^{-1} \tag{19}$$

where $Q = \underline{q}\underline{q}'/n$ and $\underline{q} = [q_1, \ldots, q_n]'$ is the vector of Press residuals, $q_i = e_i/(1 - h_{ii})$ The ordinary jackknife estimator of Var($\underline{\tilde{\beta}}$) is biased.

If the pseudovalues are centered around $\underline{\tilde{\beta}}$, instead of $\underline{\beta}_J$, then the variance estimator is as shown in (19) but without the Q term. This estimator has not been used in practice.

Hinkley (1977) suggested a weighted jackknife procedure to counter the unbalancedness of typical regression data, where unbalanced is meant in the sense that the hat values, h_{ii}, vary in size. Hinkley's variance estimator for the simple linear case can be expressed as

$$\begin{aligned}\mathrm{var}_H(\underline{\tilde{\beta}}) &= \frac{n}{n-2} (Z'Z)^{-1}(Z'\mathrm{diag}\{e_i^2\}Z)(Z'Z)^{-1} \\ &= \frac{n}{n-2} \mathrm{var}_W(\underline{\tilde{\beta}})\end{aligned} \tag{20}$$

MacKinnon and White (1985) derived (20) by applying a degrees-of-freedom adjustment to (17). Obviously, the consistency of (17) ensures the consistency of (20). Hinkley (1977) showed that $\mathrm{var}_H(\underline{\tilde{\beta}})$ is approximately unbiased and is robust against error variance heterogeneity.

Wu (1986) offered an alternative weighting scheme based on the approximate relation $E[e_i^{*2}] \cong \sigma_i^2$, where $e_i^{*2} = e_i^2/(1 - h_{ii})$ as defined earlier. The resulting covariance matrix estimator is

$$\mathrm{var}_{WU}(\underline{\tilde{\beta}}) = (Z'Z)^{-1}(Z'\mathrm{diag}[e_i^{*2}\}Z)(Z'Z)^{-1} \tag{21}$$

In certain situations, $\mathrm{var}_{WU}(\underline{\tilde{\beta}})$ is exactly unbiased, and it is biased to $0(n^{-1})$ under less restrictive conditions.

Efron and Gong (1983) outlined two methods of bootstrapping a regression equation. One has been labeled the "residual bootstrap" because OLS residuals form the basis for each bootstrap sample. Each bootstrap regression mimics the regression with the initial sample, except that the ith bootstrap observation on the dependent variable, B_i, is formed by $B_i = \underline{Z}_i'\underline{\tilde{\beta}} + b_i\{(n-2)/n\}^{0.5}$, where $\underline{\tilde{\beta}}$ is the OLS estimate and where b_i is chosen with replacement from the set of OLS residuals, $\{e_j\}_{j=1}^{n}$. For multiple linear regression models, $(n - 2)$ is replaced by $(n - t)$, where t is the number of predictors. If we let NB denote the number of bootstrap samples and $\underline{\beta}_{rb,k}$ be the estimate of $\underline{\beta}$ from the kth bootstrap sample, then the residual-bootstrap estimate of Var($\underline{\tilde{\beta}}$) is

$$\mathrm{var}_{rb}(\underline{\tilde{\beta}}) = \frac{1}{NB-1} \sum_{k=1}^{NB} (\underline{\beta}_{rb,k} - \underline{\beta}_{rb})(\underline{\beta}_{rb,k} - \underline{\beta}_{rb})' \tag{22}$$

where $\underline{\beta}_{rb}$ is the average (over the NB bootstrap samples) $\underline{\beta}_{rb,k}$. For $NB =$

∞, Var($\tilde{\beta}$) is identical to Var′, (8). In simulation studies (22) has been found to be practically indistinguishable from Var′ for even comparatively small *NB*. The residual bootstrap estimator of Var($\tilde{\beta}$) is evidently biased and inconsistent in the heteroscedastic case. Wu (1986) traces the failure of bootstrapping in this case to the nonexchangeability of the normalized residuals, $b_i\{(n - 2)/n\}^{0.5}$ in non-*iid* situations.

The second method of bootstrapping a regression forms bootstrap samples by drawing with replacement from the initial data $\{Y_i, \underline{Z}_i'\}_{i=1}^n$. Unlike the residual bootstrap variance estimator, the "simple bootstrap" estimator does not coincide with Var′, and it can be expected to perform better than the residual bootstrap under model misspecification (Efron and Gong 1983). The simple bootstrap variance estimator is similar in form to (22):

$$\mathrm{var}_{sb}(\tilde{\beta}) = \frac{1}{NB - 1} \sum_{k=1}^{NB} (\underline{\beta}_{sb,k} - \underline{\beta}_{sb})(\underline{\beta}_{sb,k} - \underline{\beta}_{sb})' \qquad (23)$$

where $\underline{\beta}_{sb}$ is the average (over the *NB* bootstrap samples) $\underline{\beta}_{sb,k}$.

An alternative estimator of variance is suggested by simple bootstrapping because each simple bootstrap regression produces an estimate, $\sigma^2_{sb,k}$, of σ^2. The alternative estimator is

$$\mathrm{var}^*_{sb}(\tilde{\beta}) = \overline{\sigma^2_{sb}}(Z'Z)^{-1} \qquad (24)$$

where $\overline{\sigma^2_{sb}}$ is the bootstrap average $\overline{\sigma^2_{sb,k}}$. This variance estimator has not appeared previously in the literature.[5]

COMPARISON OF ROBUST ESTIMATORS OF Var($\tilde{\beta}$)

A simulation experiment was designed and carried out in order to compare the performance of the robust estimators of Var($\tilde{\beta}$) described in the previous section.

SIMULATION DESIGN

The simulation consisted of the repeated generation of samples, of a fixed size, according to the specified heteroscedastic model. The model was fitted by OLS to each generated sample, and Var($\tilde{\beta}$) was estimated by the customary OLS estimator (9). In addition, $\mathrm{var}_W(\tilde{\beta})$, $\mathrm{var}_J(\tilde{\beta})$, $\mathrm{var}_H(\tilde{\beta})$, $\mathrm{var}_{WU}(\tilde{\beta})$, $\mathrm{var}_{sb}(\tilde{\beta})$, and $\mathrm{var}^*_{sb}(\tilde{\beta})$ were calculated.

The generating model was

$$Y = -0.45 + 0.002X + \epsilon \qquad (25)$$

where ϵ was an $N(0, \sigma^2 X^{\omega})$ random variable. The error terms were generated from the pseudorandom variables produced by the GGNML routine of the International Statistics and Mathematics Library (IMSL 1980). The coefficient vector $\underline{\beta} = [-0.45, 0.002]'$ was chosen to mimic common fitted values (in English units) for the combined-variable volume equation for tree merchantable volume. See, e.g., Avery and Burkhart (1983, p. 93). The value of $\underline{\beta}$ is somewhat irrelevant because the comparative performance of the alternative variance estimators will not be affected by different parameterizations of the coefficient vector.

The values of X = (dbh)^2*(height) in (25) were taken from Table 6.3 of

[5] B. Efron, pers. comm.

Avery and Burkhart (1983), thereby fixing the sample size at $n = 40$ observations. The mean value of X was approximately 30,000 in.^{2}-ft.

We anticipated that values of σ^2 and ω may affect the relative performance of competing estimators. Six values of the weight parameter were chosen for the experiment: $\omega = 0.0, 0.5, 1.0, 1.5, 2.0$, and 2.5. The null value was included because it corresponds to the homoscedastic model. The other choices, except 2.5, have appeared a number of times in the literature.

Values of σ^2 were guided by the following plan. The homoscedastic version of (25) is achieved by transformation:

$$Y^* = -0.45X_1^* + 0.002X_2^* + \epsilon^*$$

where $Y^* = YX^{-\omega/2}$, $X_1^* = X^{-\omega/2}$, $X_2^* = X^{1-\omega/2}$, $\epsilon^* = \epsilon X^{-\omega/2}$. Since $\mathrm{Var}(\epsilon^*) = \sigma^2$, then a region of width 2σ on either side of the regression line, $E[Y^*]$, should contain the bulk of the Y^* that are observed at $[X_1^*, X_2^*]$. We chose this interval half-width to be a fixed proportion, p, of $E[Y^*]$, namely

$$E[Y^*] + 2\sigma = pE[Y^*]$$

which implies that

$$\begin{aligned}\sigma &= (p - 1)E[Y^*]/2 \\ &= (p - 1)(-0.45X_1^* + 0.002X_2^*)/2 \qquad (26)\end{aligned}$$

We calculated σ according to (26) evaluated at mean X for arbitrarily chosen values of $p = 1.1, 1.2, 1.5$, and 2.0. While these values were chosen arbitrarily, they span a reasonable range for most natural resource applications.

The number of samples selected in each simulation run was set at MC = 2,000. The size of each experimental run was set at this level after preliminary testing showed that results were extremely stable for $1{,}000 \leq MC \leq 10{,}000$[6]. The level $MC = 2{,}000$ was selected to provide an extra margin of reliability to the Monte Carlo results. The number of bootstrap samples (per each Monte Carlo sample) was fixed at $NB = 200$, after preliminary testing indicated that results were fairly unchanged over the range $100 \leq NB \leq 1{,}000$.

Comparison Criteria

Our interest was focused on the estimates of the variances of $\tilde{\beta}_0$ and $\tilde{\beta}_1$ and the covariance between $\tilde{\beta}_0$ and $\tilde{\beta}_1$. The true values were calculated from (7), so that we were able to compute the bias and root mean square error of each estimator. To remove the effect of scale, we expressed bias and root mean square error relative to the true value. For $\mathrm{var}_J(\tilde{\beta}_0)$, as an example,

$$\text{bias} = \frac{E_{MC}[\mathrm{var}_J(\tilde{\beta}_0) - \mathrm{Var}(\tilde{\beta}_0)]}{\mathrm{Var}(\tilde{\beta}_0)}$$

$$\text{rmse} = \frac{\{E_{MC}[\mathrm{var}_J(\tilde{\beta}_0) - \mathrm{Var}(\tilde{\beta}_0)]^2\}^{0.5}}{\mathrm{Var}(\tilde{\beta}_0)}$$

where $E_{MC}[\cdot]$ represent the Monte Carlo average value. The covariance re-

[6] For example, for the OLS estimator of β_0, when $\sigma = 0.002$, the following rmse values were obtained: 1.11 (MC = 1,000), 1.15 (MC = 2,000), 1.09 (MC = 3,000 and 4,000), 1.11 (MC = 5,000 and MC = 10,000). For β_1, the corresponding rmse values were 0.49 (MC = 1,000, 3,000, 4,000, 5,000 and 10,000) and 0.48 (MC = 2,000).

sults were computed relative to the absolute value of $\mathrm{Cov}(\tilde{\beta}_0, \tilde{\beta}_1)$ so that the direction of the relative bias would not be obscured by the sign of the covariance, as in Wu (1986).

Low values of relative bias implies that a variance estimator, on average, can be used to construct accurate interval estimates or to test hypotheses. A variance estimator with a low relative bias but large root mean square error indicates that an individual estimate may be quite erroneous, despite the closeness of the estimator, on the average, to the true value.

Positive (negative) bias is evidence that the average error value is positive (negative), but neither the sign nor magnitude of bias provides evidence as to the relative frequency of positive and negative errors. It may be useful to know whether a positive (negative) bias results from a tendency of errors to be uniformly, or nearly so, positive (negative) or not. For this reason, we also calculated relative absolute bias, which for $\mathrm{var}_J(\tilde{\beta}_0)$, say, is

$$|\mathrm{bias}| = \frac{E_{MC}[|\mathrm{var}_J(\tilde{\beta}_0) - \mathrm{Var}(\tilde{\beta}_0)|]}{\mathrm{Var}(\tilde{\beta}_0)}$$

Values of relative bias that are close in magnitude, disregarding sign, to the relative absolute bias indicate a propensity of the estimator to uniformly under/over predict.

SIMULATION RESULTS

Without exception, varying levels of σ for a fixed value of ω had no systematic or consequential effect on the performance of any variance estimator. Therefore, for each ω level, we present results for the single σ level that corresponds to $p = 1.2$ in (26). These results appear in Table 2, parts (a)–(f).

The uniform variance ($\omega = 0$) results are shown in Table 2a. Expectedly, OLS is unbiased and has minimum rmse. The two weighted jackknife estimators were less biased and had similarly lower rmse than the ordinary jackknife, but judging by rmse, they were slightly inferior to the White estimator, var_W. The simple bootstrap estimator, var_{sb}, performs the same as the jackknife estimators, whereas var^*_{sb} mimics the OLS estimator.

ESTIMATORS OF $\mathrm{VAR}(\tilde{\beta}_0)$

The bias and rmse of the OLS estimator and var^*_{sb} are both large and both increase in magnitude with increasing ω, as witnessed in Tables 2b–2f. In most cases, the bias of these two estimators exceeds that of the other estimators by an order of magnitude. The Wu jackknife is generally least biased, but it enjoys no advantage in rmse compared to var_W, var_{sb}, and the other jackknife estimators. In contrast to the finding of Wu (1986), the Hinkley estimator, var_H, was never more biased than the OLS variance estimator, in the presence of heteroscedasticity. White's estimator, var_W, is uniformly more biased than var_H but it always has slightly lower rmse. The performance of var^*_{sb} is quite better than that of the OLS estimator, but generally it tracks the OLS estimator fairly consistently.

The absolute bias of the OLS estimator is nearly equal in magnitude to that of the bias, indicating the expected overprediction of $\mathrm{Var}(\tilde{\beta}_0)$ by the OLS estimator. No systematic under/over prediction was indicated by the other estimators, except var^*_{sb}.

The mean square error performance of all variance estimators of β_0 dete-

TABLE 2. *Bias, absolute bias, and root mean square error of variance estimators. Results are expressed relative to the true value and are based on 2,000 Monte Carlo samples at each ω.*

	Var ($\tilde{\beta}_0$)			Var ($\tilde{\beta}_1$)			Cov ($\tilde{\beta}_0$, $\tilde{\beta}_1$)		
	bias	|bias|	rmse	bias	|bias|	rmse	bias	|bias|	rmse
(a) $\omega = 0.0, \sigma = 6.0$									
OLS	0.00	0.18	0.23	0.00	0.18	0.23	0.00	0.18	0.23
var_W	−0.05	0.23	0.28	−0.07	0.26	0.33	0.06	0.24	0.30
var_J	0.03	0.24	0.31	0.07	0.29	0.39	−0.05	0.26	0.34
var_H	0.00	0.23	0.30	−0.03	0.27	0.33	0.01	0.25	0.31
var_{WU}	0.00	0.23	0.30	0.01	0.27	0.35	−0.01	0.25	0.32
var_{sb}	−0.02	0.24	0.30	−0.01	0.27	0.35	0.01	0.26	0.32
var^*_{sb}	−0.05	0.18	0.22	−0.05	0.18	0.22	0.05	0.18	0.22
(b) $\omega = 0.5, \sigma = 0.45$									
OLS	0.41	0.44	0.54	−0.17	0.22	0.27	−0.15	0.25	0.32
var_W	−0.04	0.21	0.26	−0.10	0.32	0.39	0.08	0.28	0.34
var_J	0.06	0.23	0.30	0.07	0.36	0.48	−0.07	0.31	0.42
var_H	0.01	0.22	0.27	−0.05	0.32	0.41	0.03	0.28	0.35
var_{WU}	0.02	0.22	0.28	−0.01	0.34	0.43	0.00	0.29	0.37
var_{sb}	0.00	0.23	0.29	−0.04	0.32	0.40	0.02	0.28	0.36
var^*_{sb}	0.33	0.37	0.47	−0.21	0.25	0.29	−0.09	0.22	0.29
(c) $\omega = 1.0, \sigma = 0.054$									
OLS	0.74	0.75	0.89	−0.34	0.35	0.39	−0.09	0.25	0.32
var_W	−0.07	0.24	0.30	−0.13	0.36	0.44	0.12	0.36	0.44
var_J	0.05	0.27	0.36	0.05	0.41	0.54	−0.06	0.40	0.53
var_H	−0.03	0.24	0.31	−0.08	0.37	0.46	0.07	0.36	0.45
var_{WU}	0.00	0.25	0.32	−0.03	0.38	0.48	0.02	0.37	0.48
var_{sb}	−0.02	0.25	0.32	−0.08	0.36	0.45	0.06	0.35	0.44
var^*_{sb}	0.64	0.65	0.78	−0.38	0.39	0.42	−0.03	0.23	0.29
(d) $\omega = 1.5, \sigma = 0.0026$									
OLS	0.91	0.92	1.10	−0.46	0.46	0.49	0.03	0.25	0.32
var_W	−0.09	0.31	0.40	−0.13	0.39	0.49	0.13	0.41	0.51
var_J	0.07	0.35	0.49	0.06	0.44	0.60	−0.07	0.46	0.63
var_H	−0.04	0.31	0.41	−0.08	0.40	0.50	0.08	0.41	0.53
var_{WU}	0.00	0.33	0.44	−0.02	0.41	0.54	0.02	0.43	0.57
var_{sb}	−0.04	0.31	0.40	−0.09	0.38	0.48	0.08	0.40	0.51
var^*_{sb}	0.78	0.80	0.97	−0.49	0.50	0.52	0.10	0.26	0.31
(e) $\omega = 2.0, \sigma = 0.0002$									
OLS	0.92	0.94	1.16	−0.53	0.53	0.56	0.15	0.29	0.35
var_W	−0.10	0.38	0.49	−0.13	0.42	0.52	0.13	0.45	0.57
var_J	0.08	0.44	0.62	0.07	0.47	0.65	−0.07	0.51	0.70
var_H	−0.05	0.38	0.51	−0.08	0.42	0.54	0.09	0.45	0.59
var_{WU}	0.00	0.40	0.55	−0.02	0.44	0.58	0.02	0.47	0.63
var_{sb}	−0.05	0.37	0.49	−0.10	0.41	0.52	0.10	0.44	0.56
var^*_{sb}	0.78	0.80	1.01	−0.56	0.57	0.59	0.21	0.30	0.35
(f) $\omega = 2.5, \sigma = 0.0000254$									
OLS	0.82	0.85	1.10	−0.59	0.59	0.61	0.25	0.34	0.39
var_W	−0.12	0.42	0.53	−0.14	0.42	0.52	0.15	0.45	0.57
var_J	0.07	0.48	0.67	0.05	0.47	0.64	−0.06	0.51	0.70
var_H	−0.08	0.42	0.55	−0.10	0.43	0.54	0.10	0.46	0.59
var_{WU}	−0.02	0.45	0.60	−0.04	0.44	0.58	0.04	0.48	0.63
var_{sb}	−0.08	0.41	0.53	−0.12	0.41	0.52	0.12	0.44	0.56
var^*_{sb}	0.68	0.72	0.95	−0.62	0.62	0.64	0.31	0.36	0.41

riorated with increasing ω. The deterioration of OLS was exceptionally great.

Estimators of $\mathrm{Var}(\tilde{\beta}_1)$

All estimators but the ordinary jackknife, var_J, tend to underestimate $\mathrm{Var}(\tilde{\beta}_1)$. The Wu estimator again displays least bias, but it is generally inferior to all but var_J when judged by rmse. As above, the Hinkley estimator is always less biased than the OLS estimator in all cases where $\omega \neq 0$. The OLS estimator is surprisingly competitive from a rmse standpoint, despite its severe downward bias. A comparison of bias and |bias| makes apparent the fact that OLS underestimates almost uniformly, not just on average. The modified simple bootstrap, var^*_{sb}, generally performs slightly poorer than the OLS estimator. For all estimators, the deterioration in mean square error with increasing ω is less severe than it was for $\mathrm{Var}(\beta_0)$.

Estimators of $\mathrm{Cov}(\tilde{\beta}_0, \tilde{\beta}_1)$

The OLS estimator is considerably less biased for $\mathrm{Cov}(\tilde{\beta}_0, \tilde{\beta}_1)$ than it is for the variance of either parameter. Another distinction is that it does not as consistently under/over predict at a set value of ω, and the average under/over prediction appears to trend with the value of ω. White's estimator, var_W, is generally as biased as the OLS estimator. The Wu estimator is always least biased, whereas the simple bootstrap nearly always has bias intermediate that of the OLS and Wu estimators.

The OLS estimator has uniformly smaller rmse, except for a few scattered instances when var^*_{sb} is marginally better. The ordinary jackknife has uniformly greatest rmse, which ranges from 2–7% greater than the rmse of the Wu estimator.

DISCUSSION

The gains in efficiency that are possible by modeling the error variance heterogeneity makes this method of estimation appealing when sample size is large enough to permit the parameter, ω, to be precisely estimated. We do not know the size necessary to ensure this property. It surely will depend on the range of X values spanned, the size of σ in relation to the variation in Y_i in the sample (cf. Suich and Derringer 1977), and the extent of heterogeneity, i.e., the size of ω. These factors perhaps can be used to establish guidelines that can be used to decide whether this tactic would be worthwhile in a particular setting. For very small samples, regardless of other sample attributes, it seems advisable to use one of the heteroscedasticity-robust procedures for variance estimation. The results of Freedman and Peters (1984) bolster this view. We temper this recommendation when prior information about the size of ω is available and the modeler is willing to condition estimation on this value.

Considering the added uncertainty that accompanies the estimation of the heterogeneity parameter, a prudent course of action regardless of sample size may be to use OLS with one of the alternative variance estimators. But there are at least two drawbacks to this strategy. First, one's estimate of σ may be very substantially distorted. Second, substantial gains in efficiency may be sacrificed, thereby reducing the power of hypothesis tests and augmenting the width of confidence intervals and prediction intervals. An indication of the inefficiency of OLS is provided by the sample data used in the simulation experiment. For the $\omega = 2$ model, the ratio of (true) OLS to GLS standard errors are 4.47 and 1.79 for the estimators of β_0 and β_1, respectively. The ratio of generalized variances is 23.7. In finite sample situations, one can rarely expect to model the heterogeneity and realize the efficiency that accompanies GLS. Nonetheless, one might realize a nonnegligible advantage by so doing.

If the prudent course of action is followed, the results in Table 2 are encouraging. With the exception of var^*_{sb}, the robust variance estimators exhibit substantially less bias and lower mean square error in the case of $\text{Var}(\beta_0)$ and substantially less bias and nearly equivalent mean square error in the case of $\text{Var}(\beta_1)$. Another advantage to the use of the alternative variance estimators is the comparative freedom from other assumptions about error distribution. In particular, none of the jackknife or bootstrap alternative estimators rely on an assumption of normally distributed errors. Normality may not always be a tenable assumption, yet it is obviously integral to the ML estimators. Normality, too, plays an important role in the derivation of many of the asymptotic properties of FGLS estimators, although some work has been done to develop estimators that are robust to departures from normality (Roshwalb 1987).

Table 1 results suggest that ML estimates of ω exceed those based on squared residuals or transforms of squared residuals. Our results give no indication of which estimator is preferable, i.e., which is less prone to systematic and stochastic error. An investigation into this matter is planned.

Table 2 results indicate that OLS estimators are competitive for some variance terms when judged by mean square error but not when judged by bias. Researchers will need to decide which of these criteria is the more important for valid inference and for accurate prediction. An overall measure of goodness is needed, and working by analogy to generalized variance, the determinant of the mean square error matrix is suggested. It is not clear to us, however, whether this is a meaningful measure.

We conjecture that many researchers will decide to model heteroscedasticity, ignore heteroscedasticity, or use robust alternatives (for variance estimation) for nonstatistical reasons, at least until the theoretical and practical consequences of these courses of action become better determined. For those who seek to model heteroscedasticity, we suggest FGLS or ML as a desirable alternative to the "interval estimation" approach, irrespective of sample size. When bias in variance estimators of OLS estimators is of primary concern, we suggest Wu's jackknife estimator. When mean square error of variance estimators is of primary importance, OLS is as good as any of the alternative variance estimators, except for the intercept parameter.

Finally, we conjecture that variance heterogeneity of the form $\sigma_i^2 \propto g(E[Y_i])$ may be more tenable than the multiplicative heteroscedasticity that has traditionally been assumed in forestry. If it is at least a viable alternative, then the rich family of Box-Cox transformations become available tools. Other alternatives based on squared residuals are available, also.

LITERATURE CITED

AMEMIYA, T. 1977. A note on a heteroscedastic model. J. Econ. 6:365–371.

AVERY, T. E., and H. E. BURKHART. 1983. Forest measurements. Ed. 3. McGraw-Hill, New York. 331 p.

BELSLEY, D. A., E. KUH, and R. E. WELSCH. 1980. Regression diagnostics. Wiley, New York. 292 p.

BLOCH, D. A., and L. E. MOSES. 1988. Nonoptimally weighted least squares. Am. Statist. 42:50–53.

BLOOMFIELD, P., and G. S. WATSON. 1975. The inefficiency of least squares. Biometrika 62:121–128.

BOX, G. E. P., and D. E. COX. 1964. The analysis of transformations. J. Roy. Statist. Soc., Series B, 26:211–252.

————, and W. J. HILL. 1974. Correcting inhomogeneity of variance with power transformation weighting. Technometrics 16:385–389.

————, and P. W. TIDWELL. 1962. Transformation of the independent variables. Technometrics 4:531–550.

BURKHART, H. E., R. C. PARKER, and R. G. ODERWALD. 1972. Yields for natural stands of loblolly pine. School of For. and Wildl. Resour. Publ. FWS-2-72, Virginia Polytechnic Institute and State University. 63 p.

CARROLL, R. J., and D. RUPPERT. 1982. Robust estimation in heteroscedastic linear models. Ann. Statist. 10:429–441.

————. 1985. Transformations in regression: A robust analysis. Technometrics 27:1–12.

————. 1986. Discussion of C. F. J. Wu. Ann. Statist. 14:1298–1301.

CHEW, V. 1970. Covariance matrix estimation in linear models. J. Am. Statist. Assoc. 65:173–181.

CRAGG, J. G. 1983. More efficient estimation in the presence of heteroscedasticity of unknown form. Econometrica 51:751–763.

CUNIA, T. 1964. Weighted least squares method and construction of volume tables. For. Sci. 10:180–191.

————. 1965. Some theory on the reliability of volume estimates in a forest inventory sample. For. Sci. 11:115–128.

DAVIDIAN, M., and R. J. CARROLL. 1987. Variance function estimation. J. Am. Statist. Assoc. 82:1079–1091.

DEATON, M. L., M. R. REYNOLDS, and R. H. MYERS. 1983. Estimation and hypothesis testing in regression in the presence of nonhomogeneous error variances. Commun. Statist. —Simul. Comput. 12:45–66.

DON, F. J. H., and J. R. MAGNUS. 1980. On the unbiasedness of iterated GLS estimators. Commun. Statist.—Theor. Meth. A9:519–527.

DRAPER, N. R., and H. SMITH. 1981. Applied regression analysis. Wiley, New York. 709 p.

EFRON, B. 1982. The jackknife, the bootstrap, and other resampling plans. SIAM, Philadelphia. 92 p.

————, and G. GONG. 1983. A leisurely look at the jackknife, the bootstrap, and cross-validation. Am. Statist. 37:36–48.

FORD, E. D. 1975. Competition and stand structure in some even-aged plant monocultures. J. Ecol. 58:272–296.

FREEDMAN, D. A., and S. C. PETERS. 1984. Bootstrapping a regression equation: Some empirical results. J. Am. Statist. Assoc. 79:97–106.

FULLER, W. A., and J. N. K. RAO. 1978. Estimation for a linear regression model with unknown diagonal covariance matrix. Ann. Statist. 6:1149–1158.

GEDNEY, D. R., and F. A. JOHNSON. 1959. Weighting factors for computing the relation between tree volume and D.B.H. in the Pacific Northwest. U.S.D.A. For. Serv. Res. Note PNW 174. 5 p.

GOLDFIELD, S. M., and R. E. QUANDT. 1972. Nonlinear methods in econometrics. North Holland, Amsterdam.

GREGOIRE, T. G. 1984. The jackknife: An introduction with applications in forestry data analysis. Can. J. For. Res. 14:493–497.

————, H. T. VALENTINE, and G. M. FURNIVAL. 1986. Estimation of bole volume by importance sampling. Can. J. For. Res. 16:554–557.

HARTLEY, H. O., and K. S. E. JAYATILLAKE. 1973. Estimation for linear models with unequal variances. J. Am. Statist. Assoc. 68:189–192.

HARVEY, A. C. 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica 44:461–465.

HILDRETH, C., and J. P. HOUCK. 1968. Some estimators for a linear model with random coefficients. J. Am. Statist. Assoc. 63:584–595.

HINKLEY, D. V. 1977. Jackknifing in unbalanced situations. Technometrics 19:285–292.

HORN, S. D., R. A. HORN, and D. B. DUNCAN. 1975. Estimating heteroscedastic variances in linear models. J. Am. Statist. Assoc. 70:380–385.

IMSL. 1980. International Mathematical and Statistical Libraries. Ed. 9. Author, Houston, TX.

JACQUES, J. A., F. J. MATHER, and C. R. CRAWFORD. 1968. Linear regression with non-constant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. J. Am. Statist. Assoc. 24:607–626.

JOBSON, J. D., and W. A. FULLER. 1980. Least squares estimation when the covariance matrix and parameter vector are functionally related. J. Am. Statist. Assoc. 75:176–181.

KAKWANI, N. C. 1967. The unbiasedness of Zellner's seemingly unrelated regression equations estimators. J. Am. Statist. Assoc. 2:141–142.

KMENTA, J. 1986. Elements of econometrics. Ed. 2. MacMillan, New York. 786 p.

MACKINNON, J. G., and H. WHITE. 1985. Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties. J. Econ. 29:305–325.

MAGNUS, J. R. 1978. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. J. Econ. 7:281–312.

MCCLURE, J. P., and R. L. Czaplewski. 1987. High order regression models for regional volume equations. U.S.D.A. For. Serv. Tech. Rep. NE-GTR-117.

————, H. T. SCHREUDER, and R. L. WILSON. 1983. A comparison of several volume equations for loblolly pine and white oak. U.S.D.A. For. Serv. Res. Pap. SE-240.

MENG, C. H., and W. Y. TSAI. 1986. Selection of weights for a weighted regression of tree volume. Can. J. For. Res. 16:671–673.

MYERS, R. H. 1986. Classical and modern regression with applications. Duxbury Press, Boston. 359 p.

RAO, C. R. 1970. Estimation of heteroscedastic variances in linear models. J. Am. Statist. Assoc. 65:161–172.

ROSHWALB, A. 1987. The estimation of heteroscedastic linear model using inventory data. Paper presented at Am. Statist. Assoc. annual meeting, San Francisco.

ROTHENBERG, T. J. 1984. Approximate normality of generalized least squares estimates. Econometrica 52:811–825.

SATHE, S. T., and H. D. VINOD. 1974. Bounds on the variance of regression coefficients due to heteroscedastic or autoregressive errors. Econometrica 42:333–340.

SCHREUDER, H. T., and W. T. SWANK. 1973. A comparison of several statistical methods in forest biomass and surface area estimation. Misc. Publ. 132, Agric. Exp. Stn., Univ. of Maine, Orono.

SCHUMACHER, F. X., and R. A. CHAPMAN. 1942. Sampling methods in forest and range management. Duke University, School of Forestry, Bull. 7. Durham, NC. 214 p.

SEBER, G. A. F. 1977. Linear regression analysis. Wiley, New York. 465 p.

SPURR, S. H. 1952. Forest inventory. Wiley, New York. 214 p.

SUICH, R., and G. C. DERRINGER. 1977. Is the regression equation adequate?—one criterion. Technometrics 19:213–216.

SWINDEL, B. F. 1968. On the bias of some least-squares estimators of variance in a general linear model. Biometrika 55:313–316.

TAYLOR, W. E. 1978. The heteroscedastic linear model: Exact finite sample results. Econometrica 46:663–676.

THEIL, H. 1971. Principles of econometrics. Wiley, New York. 736 p.

WHITE, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838.

Wu, C. F. J. 1986. Jackknifing, bootstrapping, and other resampling methods in regression analysis. Ann. Statist. 14:1261–1295.

APPENDIX

The variance estimator of $\underline{\beta}_G^*$ is

$$\mathrm{var}(\underline{\beta}_G^*) = \sigma^{2*}(Z'V^{*-1}Z)^{-1} \tag{A1}$$

where

$$\begin{aligned}\sigma^{2*} &= (\underline{Y} - Z\underline{\beta}_G^*)'V^{*-1}(\underline{Y} - Z\underline{\beta}_G^*)/(n-2)\\ &= \underline{\epsilon}'(I_n - Z(Z'V^{*-1}Z)^{-1}Z'V^{*-1})'V^{*-1}\\ &\quad (I_n - Z(Z'V^{*-1}Z)^{-1}Z'V^{*-1})\underline{\epsilon}/(n-2)\\ &= \underline{\epsilon}'(I_n - Z(Z'V^{*-1}Z)^{-1}Z'V^{*-1})'V^{*-1}\underline{\epsilon}\end{aligned} \tag{A2}$$

The bias of σ^{2*} is

$$E[\sigma^{2*}] - \sigma^2 = \sigma^2\mathrm{trace}[\{(I_n - Z(Z'V^{*-1}Z)^{-1}Z'V^{*-1})'V^{*-1}V\}/(n-2) - 1] \tag{A3}$$

The variance estimator of $\tilde{\underline{\beta}}_G$ is

$$\mathrm{var}(\tilde{\underline{\beta}}_G) = \sigma_G^2(Z'\tilde{V}^{-1}Z)^{-1} \tag{A4}$$

where

$$\begin{aligned}\sigma_G^2 &= (\underline{Y} - Z\tilde{\underline{\beta}}_G)'\tilde{V}^{-1}(\underline{Y} - Z\tilde{\underline{\beta}}_G)/(n-2)\\ &= \underline{\epsilon}'(I_n - Z(Z'\tilde{V}^{-1}Z)^{-1}Z'\tilde{V}^{-1})'\tilde{V}^{-1}\underline{\epsilon}\end{aligned} \tag{A5}$$

and where $\tilde{V}$ is constructed with the prior FGLS estimate of ω. The bias of (A5) is as shown in (A3) but with $\tilde{V}$ in place of V^*. For the ML estimators of $\underline{\beta}$,

$$\mathrm{var}(\beta_0) = \frac{\sigma_{ML}^2}{D}\sum_{i=1}^{n} X_i^{2-\omega}$$

$$\mathrm{var}(\beta_1) = \frac{\sigma_{ML}^2}{D}\sum_{i=1}^{n} X_i^{-\omega}$$

where

$$D = \left(\sum_{i=1}^{n} X_i^{2-\omega}\right)\left(\sum_{i=1}^{n} X_i^{-\omega}\right) - \left(\sum_{i=1}^{n} X_i^{1-\omega}\right)^2$$