VALUING FOREST LAND USING REGRESSION:
A PROBLEM IN MODEL SPECIFICATION AND
GOODNESS OF FIT

B. Bruce Bare and Guillermo A. Mendoza,
Professor, College of Forest Resources and Center for
Quantitative Science in Forestry, Fisheries and Wildlife,
University of Washington, Seattle, WA 98195 and
Assistant Professor, Department of Forestry,
University of Illinois, Urbana, IL 61801

SUMMARY

Bare forest land in western Washington is valued for property tax
purposes based on its current forest use value. Several years ago,
a regression model was proposed for estimating this value using
actual forest land transactions evidence. Because of a dearth of
market evidence involving the sale of only bare forest acres, it
was necessary to use a model which had the total sales price of the
parcel as the dependent variable with independent variables
consisting of the number of acres in the sale stocked with trees
of different ages and conditions. These independent variables
represented the components of value to which the sales price was
to be allocated. Site quality, access, and topography were
considered as determinants of this value allocation, but not as
value components. The regression coefficients represent the per
acre value of each of the value components. A variety of different
model forms were tested using ordinary and weighted least squares.
However, the resulting parameter estimates and goodness of fit
indicators were vastly different - even for comparable models --
causing some analysts to recommend that the regression approach be
abandoned. This paper explores the assumptions underlying these
different model forms, the appropriateness of the goodness of fit
indicators, and steps taken to check on the degree of
multicollinearity present in the data.

INTRODUCTION

Washington's 1971 Forest Tax Law exempts standing timber from the
annual ad valorem tax, but substitutes a 5 percent yield tax at
time of harvest. Forest land, however, remains subject to the
annual ad valorem property tax. Since 1972, three forest land
valuation methodologies have been used in an effort to estimate
the value of bare forest land: (a) abstraction, (b) multiple
regression, and (c) various income approaches. All methodologies
aim at estimating the value of bare forest land assuming that the
highest and best use of the land is for the continued production
of forest crops. Presently, about 6.8 million acres are subject to the 1971 Forest Tax Law.

Given the legal requirement that bare land values must reflect the use of the land solely for the growing and harvesting of timber, and that the 1971 Forest Tax Law requires a valuation based on true and fair value -- interpreted in Washington to mean fair market value -- the decision was made to use actual sales of forest land as evidence of the land value. Problems arise because very few forest land sales involve only bare land. In fact, the majority of sales involve land plus timber (mature and immature) values. In order to increase the number of transactions in the data base, the decision was made to include sales involving these additional value elements. This then leads to the problem of segregating the value of bare land from the total purchase price. The analytical process used in separating the value elements became the focus of an intense legal battle which erupted in 1977, with the three methodologies mentioned above playing pivotal roles in this debate.

From 1972 until 1978, abstraction was the method used by the Washington State Department of Revenue (DOR) to make its determination of the value of bare forest land in the State of Washington. Bare (1978, 1983) provides a detailed discussion of this procedure along with a list of criticisms. This process produces estimates of each value element in a sequential fashion and is similar, in concept, to solving a system of linear equations sequentially and not simultaneously. With repeated substitution, an approximate, feasible solution is possible to derive. However, the judgement of the analyst plays a crucial role in selecting the final value estimates. Because preconceived biases can easily distort this decision, abstraction is not reliable as a land value estimation methodology and is no longer used in Washington (Bare and McKetta, 1977).

REGRESSION MODEL DEVELOPMENT

Starting in 1977, the DOR began to experiment with the use of multiple regression as a possible land valuation methodology. The same market evidence and model that were used in abstraction were retained. However, the simultaneous principle of least squares replaced the sequential estimation process of abstraction. The model used is shown below as eq. (1):

\[ Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_i \]  

where,

\[ Y_i = \text{Sale price of } i^{th} \text{ forest land sale net of merchantable timber value.} \]

\[ X_1 = \text{Number of acres of bare land on } i^{th} \text{ sale.} \]

\[ X_2 = \text{Number of acres of conifer reproduction on } i^{th} \text{ sale.} \]

\[ X_3 = \text{Number of acres of immature conifer on } i^{th} \text{ sale.} \]

\[ X_4 = \text{Number of acres of hardwood reproduction on } i^{th} \text{ sale.} \]

\[ X_5 = \text{Number of acres of immature hardwood on } i^{th} \text{ sale.} \]

\[ \epsilon_i = \text{Error or deviation between actual net sale price and predicted net sale price.} \]

\[ \beta_0 = \text{Per acre value of bare forest land.} \]

\[ \beta_1 = \text{Per acre value of conifer reproduction (1-5 inches diameter) and land.} \]

\[ \beta_2 = \text{Per acre value of immature conifer (5-8 inches diameter) and land.} \]

\[ \beta_3 = \text{Per acre value of hardwood reproduction (1-5 inches diameter) and land.} \]

\[ \beta_4 = \text{Per acre value of immature hardwood (5-8 inches diameter) and land.} \]

Each sale is also characterized by its site productivity, access, and topography. However, in eq. (1) all sales are considered in the aggregate to determine average values for each element. Ex post facto adjustments are subsequently made for other classes of land.

For theoretical reasons, the model specified in eq. (1) does not contain an intercept term and, as discussed later, this leads to problems of interpretation when the goodness of fit of the regression model is considered. Furthermore, the variance in net sale price increases with increasing sale size. Thus, a transformation or weighting is needed to obtain the minimum variance linear unbiased estimate of each value element. Lastly, the model contains a theoretical restriction: \( \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 = \tau \), where \( \tau \) is as defined below. Thus, ill conditioning of the \( X'X \) matrix is a potential concern.

If adequate numbers of sales involving only bare land are available, a much simpler model to consider is:

\[ Z_i = f(S_{i1}, A_{i1}, T_{i1}) \]  

where,

\[ Z_i = \text{Per acre value of } i^{th} \text{ bare land sale.} \]

\[ S_{i1} = \text{Site index measured at age 100.} \]

\[ A_{i1} = \text{Indicator variable for access-topography (1=difficult, 2=average, 3=favorable).} \]

\[ T_{i1} = \text{Total acres on } i^{th} \text{ sale.} \]

Eq. (2) directly estimates the variable of interest (i.e., the bare land value) as a function of site quality, access-topography, and sale size. Here, the values of the regression coefficients are less important than in eq. (1). However, eq. (2) was not used in this study due to a lack of adequate sales involving only bare land.
Regression analyses were carried out for the years 1975 - 1981, but the year 1977 is used here for demonstration purposes. This was also the year over which the aforementioned legal battle was centered. The following summary statistics reveal the character of the 159 sales used in the analysis of the 1977 data for western Washington: (1) 49,851 acres of forest land with a total net sale price of $8,590,602 or $172.33/acre were analyzed, (2) average net sale price of $54,030/sale with a standard deviation of $125,317/sale, (3) average sale size of 313.53 acres with a range of 20 - 9,035 acres, (4) 2,865 acres involving 33 sales included only bare land and sold for $451,418 or $157.56/acre, (5) sales involving only bare land ranged in size from 20 - 459 acres, (6) acreage summaries for the five value elements shown in eq. (1) were BL - 17,097, CR - 15,921, CI - 3,474, HR - 1,961, and HI - 11,398, (7) average site index was 144.5, and (8) average access-topography class was 2.54.

Using eq. (1), the first task is to identify a suitable weight to use to produce the desired homogeneous variance. An examination of a plot of the residuals for the 1977 sales using eq. (1) demonstrated that the variance increased with increasing sale size. To correct for this, weighted least squares was used where the weight \(w_i\) was set equal to \(1/TA_i\). Equivalently, eq. (1) can be transformed by dividing both sides by \(TA_i\) -- thus allowing use of ordinary least squares. When this is done, we obtain eq. (2):

\[
Y_i/TA_i = \beta_1 BL_i/TA_i + \beta_2 CR_i/TA_i + \beta_3 CI_i/TA_i + \beta_4 HR_i/TA_i + \beta_5 HI_i/TA_i + \epsilon_i/TA_i
\]

(3)

An examination of residuals for this model shows that the variance is now stabilized. Table 1 contains the parameter estimates and summary statistics using this model for 1977 western Washington.

\(R^2\) as shown above is the ratio of SSR/TSS, where both are uncorrected for the mean and degrees of freedom.

It can be seen that eq. (3) takes the form of a mixture model, wherein the X’s sum to one for all data points (Marquardt and Snee, 1974). Specifically, eq. (3) is written in Scheffé canonical form and contains no intercept. The model can be re-written in slack variable form to include an intercept term because:

\[
BL_i/TA_i + CR_i/TA_i + CI_i/TA_i + HR_i/TA_i + HI_i/TA_i = 1 \text{ for all } i
\]

(4)

Thus, solving for \(BL_i/TA_i\), eq. (3) becomes:

\[
Y_i/TA_i = \beta_1 + \beta_2 CR_i/TA_i + \beta_3 CI_i/TA_i + \beta_4 HR_i/TA_i + \beta_5 HI_i/TA_i + \epsilon_i/TA_i
\]

(5)

Where,

\(\beta_1\) = Per acre value of conifer reproduction exclusive of land.

\(\beta_2\) = Per acre value of immature conifer exclusive of land.

\(\beta_3\) = Per acre value of hardwood reproduction exclusive of land.

\(\beta_4\) = Per acre value of immature hardwood exclusive of land.

\(\beta_5\), the slack variable, remains as defined for eqs. (1) and (2).

After applying ordinary least squares to eq. (5) the results shown in Table 2 are obtained.

---

**Table 1. Regression Results for western Washington using Eq. (3).**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>147.17</td>
<td>15.59</td>
<td>9.44**</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>219.84</td>
<td>21.81</td>
<td>10.08**</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>240.28</td>
<td>33.19</td>
<td>7.24**</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>204.98</td>
<td>49.11</td>
<td>4.17**</td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>218.47</td>
<td>23.40</td>
<td>9.34**</td>
</tr>
</tbody>
</table>

\(R^2\) (uncorrected for mean) = .758

\(F = 96.38\) (significance level: 0.00)

Max VIF = 1.049

Determinant of correlation matrix = .9252

**Table 2. Regression Results for western Washington using Eq. (5).**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>147.17</td>
<td>15.59</td>
<td>9.44**</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>72.68</td>
<td>28.77</td>
<td>2.53*</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>93.12</td>
<td>37.35</td>
<td>2.49*</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>57.82</td>
<td>52.23</td>
<td>1.11</td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>71.30</td>
<td>29.47</td>
<td>2.42*</td>
</tr>
</tbody>
</table>

\(R^2\) (corrected for mean) = .072

\(F = 2.99\) (significance level: .021)

Max VIF = 1.237

Determinant of correlation matrix = .78212
value. The t statistics in Table 2 show that these four immature timber classes do not explain as much of the remaining variation in comparison with eq. (3). However, all variables remain in the model for allocation purposes.

Hahn (1977), Marquardt and Snee (1974), Kvålseth (1985), and Willett and Singer (1988) discuss aspects of the use and interpretation of $R^2$ in non-intercept, mixture models involving weighted least squares regression. Ignoring the fact that weighted least squares is used, Marquardt and Snee (1974) recommend that $R^2$ as computed in Table 2 for the slack variable mixture model, is also the correct goodness of fit statistic for the canonical Schäffé form. To use the $R^2$ value as shown in Table 1 leads to inflated goodness of fit statistics. Kvålseth (1985) also advocates use of $R^2$ as shown in Table 2, and Willett and Singer (1988) extend his work to weighted least squares by recommending that the appropriate $R^2$ statistic is one based on the untransformed dependent variable (i.e. net sale price) and not net sale price per acre.

During the court hearing, analysts introduced several variations of $R^2$ adjusted for degrees of freedom and the mean of the dependent variable. For eq. (5), where a correction for the mean net sale price per acre has already been made, $R^2$ is simple and straightforward. The adjusted $R^2 = 1 - ((n-1)\times SSE)/((n-k)\times CTSS)$, where $k$ = the number of variables (including the intercept) and $n$ = the number of observations. Therefore, for eq. (5) the adjusted $R^2 = .048$, with $k = 5$ and $n = 159$. However, this is in terms of net sale price per acre (i.e. the transformed dependent variable). For eq. (3), where regression is forced through the origin, the calculation of an adjusted $R^2$ is not as straightforward. Some regression programs compute the adjusted $R^2$ by using the TSS (uncorrected for the mean) in place of the CTSS in the above formula, leading to an adjusted $R^2 = .752$. Other programs use the TSS (uncorrected for the mean) but substitute n in place of $(n - 1)$, producing an adjusted $R^2 = .750$. Following Marquardt and Snee (1974), an adjusted $R^2 = .048$ is obtained for both mixture models shown in eqs. (3) and (5). Kvålseth (1985) recommends that the adjusted $R^2$ be computed by substituting n in place of $(n - 1)$, producing an adjusted $R^2 = .042$. Again, these are all in terms of the transformed dependent variable and not the original metric.

WEIGHTED REGRESSION

Rather than transforming eq. (1) to correct for the unequal variation in net sale price, weighted least squares regression can be used in conjunction with eq. (1). Here, the weight ($w_i$) is set equal to $1/Ta_i$. This model is equivalent to that shown in eq. (3),

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Deg. Freedom</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td></td>
<td>5 591927 (SSR)</td>
<td>1183985</td>
</tr>
<tr>
<td>Due to intercept</td>
<td></td>
<td>1 5772922 (SSR)</td>
<td>5772922</td>
</tr>
<tr>
<td>Due to variables</td>
<td></td>
<td>4 170005 (SSR)</td>
<td>36751</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>154 189148 (SSR)</td>
<td>12284</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>159 761165 (TSS)</td>
<td></td>
</tr>
</tbody>
</table>
but requires that a weighted least squares routine be used when
estimating the coefficients.

While weighted regression produces estimates of the regression
coefficients which are identical to those produced by ordinary
least squares operating on the equivalent transformed model, it is
sometimes more relevant to investigate the performance of the
regression model in terms of the original (untransformed) dependent
variable. During the 1977 court case, this also became a point of
disagreement. When undertaking weighted regression, many regression
routines actually transform the data and produce an ANOVA table
expressed in terms of the transformed variables. This is acceptable
practice and leads to $R^2$ and standard error of the estimate
statistics consistent with the transformed model. However, it is
also appropriate to calculate these statistics in terms of the
original variables. This leads to what Willeit and Singr (1988)
label a "pseudo weighted least squares $R^2$" statistic. As an
example, when we apply the regression estimates from Table 1 to the
original (untransformed) variables shown in eq. (1), the results
shown in Table 3 are obtained.

Table 3. Weighted Regression Results for western Washington using
Eq. (1).

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Deg. Freedom</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>5</td>
<td>7.134499 E12 (SSR)</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>154</td>
<td>6.547942 E11 (SSE)</td>
<td>4.251911 E9</td>
</tr>
<tr>
<td>Total</td>
<td>159</td>
<td>7.789244 E12 (TSS)</td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ (uncorrected for mean) = .916

Standard error of estimate $= \sqrt{\frac{SSR}{df}} = 6520.68$

$R^2$ (adjusted for mean and degrees of freedom) = .908

The $R^2 = .916$ is comparable to the $R^2 = .758$ shown in Table 1, and
demonstrates the significance of reporting goodness of fit in terms
of the original metric and not the transformed metric. Similarly,
the adjusted $R^2 = .908$, computed in the original metric using the
formula recommended by Marquardt and Snee (1974), is comparable to
the adjusted $R^2 = .048$ for the transformed model. For comparison,
the $R^2$ for eq. (1) when ordinary least squares is used is .981
(unadjusted for the mean or the number of independent variables).
Lastly, the pseudo $R^2$ recommended by Willet and Singer (1988), as
computed in the original metric, results in an $R^2 = .911$. This is
adjusted for the mean but not the number of independent variables.
The $R^2$ values computed in the original metric indicate that eq. (1)
is performing satisfactorily and is a reliable model for bare land
value estimation.

MODEL SPECIFICATION

Apart from concerns over goodness of fit, critics alleged that the
regression estimates shown in Tables 1 and 2 did not produce
reasonable value relationships. That is, the progression in value
from bare land to land stocked with immature timber did not stand
up when judged against other economic evidence. Further, it was
argued that this occurred because too much of the net sale price
was allocated to bare land and not enough to immature timber
values. Bare (1983) discusses this criticism in detail, and
concludes that these criticisms are flawed in that they assume
that the data base should reflect only value relationships appropriate
to sales where the current immature timber stands are identical in
all respects to the future stands to be grown on the same acre.
Since this is rarely the case, one would not expect to observe the
progression in value espoused by the critics.

Following this line of reasoning, Dowdle (1978) introduced a
constrained regression model into the argument. This model is based
on the premise that the value of bare land plus immature timber
can be represented as $\beta_1 = \beta_0 e^{\beta_1}$, where $\beta_0$ is the value of bare land
plus regeneration costs and $\beta_1$ is the value of bare land and
immature timber t-years later compounded continuously at i percent
interest. In using this model, the DOR accepted Dowdle's (1978)
assumptions that: (1) $\beta_1 = \beta_0$, (2) $\beta_2 = \beta_1 e^{\beta_1}$, (3) $\beta_3 = \beta_1 e^{\beta_1}$,
(4) $\beta_4 = \beta_2 + \$50$, and (5) $\beta_5 = \beta_1$. Further, assumptions were that $i = 6.25$
percent, $t_1 = 16.4$ years, and $t_2 = 30.4$ years. Upon
substitution of the above terms into eq. (1), the following model
was obtained:

$$Y_i = \$50Hi_1 = \beta_1 (BL_i + CR_i e^{\beta_1} + CI_i e^{\beta_1} + HR_i + HH_i) + \epsilon_i$$

where, all terms are as defined for eq. (1). Apart from the obvious
value relationship constraints embedded in this model, neither the
DOR nor Dowdle (1978) examined the need to correct for unequal
variance about the new dependent variable in eq. (6). Further, by
equating $\beta_1 = \beta_0$, both parties assumed that all sales in the data
base involved natural regeneration at a cost of zero dollars per
acre.

As for eq. (1), an examination of the residuals for the 1977 sales
using eq. (6) again demonstrates that the variance increased with
increasing sale size. Therefore, both sides of eq. (6) were divided
by $T_A$, to correct for this problem. After this transformation, the
results shown in Table 4 were obtained.

The average net sale price adjusted by the $50 per acre immature
hardwood value is $180.76 and the standard error of the estimate
for eq. (6) is $139.58 -- leading to a coefficient of variation of
77.2 percent. Interestingly, if the adjusted $R^2$ is computed by
adjusting both degrees of freedom and the mean of the transformed dependent variable, the figure of .00 is obtained. Thus, eq. (6) is judged statistically inferior to either eq. (3) or (5).

Apart from these statistical arguments, the valuation assumptions embedded in eq. (6) are questionable. Of paramount concern is the assumption that $\beta_1 = \beta_2$. Both the DOR and Dowlle interpreted the $\beta_1$ regression estimate as a bare land value. Yet, it is clear that unless all sales in the 1977 data base involve natural regeneration, some of the value ascribed to bare land should be deducted as a regeneration cost.

### Table 4. Regression Results for western Washington using Eq. (6) after Transformation.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>65.36</td>
<td>4.48</td>
<td>14.58**</td>
</tr>
</tbody>
</table>

$R^2$ (uncorrected for mean) = .573

$F = 212.43$ (significance level: .000)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Deg. Freedom</th>
<th>SS (SSR)</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>4138662</td>
<td>4138662</td>
</tr>
<tr>
<td>Error</td>
<td>158</td>
<td>3078183</td>
<td>19482</td>
</tr>
<tr>
<td>Total</td>
<td>159</td>
<td>7216845</td>
<td>(TSS)</td>
</tr>
</tbody>
</table>

Secondly, the value relationship constraints rest on a flawed logic. For example, suppose we examine a sale which involves only the conifer reproduction value element involving a 16.4 year old naturally regenerated stand selling for a net sale price of $250 per acre. Using Dowdle's model, the underlying bare land value is computed as $\beta_1 = \beta_2 e^{1/4}$, or $\beta_1 = 89.70$ per acre. To understand why this logic is incorrect consider the following information. Suppose the buyer of this sale intends to clear the current 16.4 year old stand at a cost of $75 per acre and establish a new plantation. Further, assume that the economic bare land value under this regime is $325 per acre. Given this information, the buyer will offer $250 per acre for the sale (i.e. $325 - 75$). Dowdle's model, however, will estimate the bare land value as $89.70 per acre, when in fact it is $325. This logical flaw of assuming that an observed net sale price can always be discounted at the prevailing rate of interest to obtain the underlying bare land value casts serious doubt on the utility of eq. (6). And, it helps explain the large discrepancy between the estimates of $\beta_1$ produced by eqs. (3), (5), and (6).

Another embellishment to eq. (5) is to incorporate site quality access, and topography directly into the model, rather than treating these value modifiers as exogenous variables. Eqs. (3) or (5) are only useful for computing a bare land value representative of the average acre contained in the data base. However, the DOR responsible for determining a bare land value schedule for a classes of land. The proposed model is shown below.

$$Y_i/T_A = \beta_1^r f(S_i, A_i) + \beta_2^* C_i/T_A + \beta_3^* C_i/T_A + \beta_4^* H_i/T_A + \beta_5^* H_i/T_A$$

The $\beta_i f(S_i, A_i)$ term quantifies the relationship of value acro different site quality and access-topography classes. Due to lack of a suitable theoretical function, two alternative forms we tested. In the first case, (see Bare and Harn (1981), shown in the left half of Table 5, $\beta_1$ represents the value of bare land per unit of MAI, access, and topography for a given site class, while in the second case $\beta_i$ represents the value of bare land per unit of SI index, access, and topography. The results using each of the functional forms are shown below in Table 5.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Deg. Freedom</th>
<th>SS (SSR)</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>4138662</td>
<td>4138662</td>
</tr>
<tr>
<td>Error</td>
<td>158</td>
<td>3078183</td>
<td>19482</td>
</tr>
<tr>
<td>Total</td>
<td>159</td>
<td>7216845</td>
<td>(TSS)</td>
</tr>
</tbody>
</table>

The standard error of the estimate for the two functional forms $\$111.87$ and $\$113.49$ per acre, respectively. Thus, the first fo appears to be slightly superior to the second.

As previously mentioned, the DOR used an ex post facto calculati to adjust its derived aggregate bare land value for other class
of site productivity, access, and topography. In 1977, three site classes and three access-topography classes were used. The derived land value was adjusted to reflect the three site classes in proportion to its productive potential as reflected by board foot yields from a normal yield table at an assumed rotation age of 60 years. Adjustments for access-topography reflected the present value of assumed hauling and logging costs at rotation-end. Both abstraction and regression estimators were subjected to identical adjustments. For 1977, these two sets of contested values are compared with those produced by the two functional forms in Table 6.

<table>
<thead>
<tr>
<th>Site</th>
<th>Access-Topo. Abstract.</th>
<th>DOR Regres.</th>
<th>f₁²</th>
<th>f₂²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>111</td>
<td>203</td>
<td>169</td>
<td>170</td>
</tr>
<tr>
<td>Average</td>
<td>93</td>
<td>185</td>
<td>112</td>
<td>113</td>
</tr>
<tr>
<td>Poor</td>
<td>62</td>
<td>154</td>
<td>56</td>
<td>57</td>
</tr>
<tr>
<td>Average</td>
<td>79</td>
<td>145</td>
<td>130</td>
<td>140</td>
</tr>
<tr>
<td>Poor</td>
<td>44</td>
<td>80</td>
<td>95</td>
<td>110</td>
</tr>
<tr>
<td>Average</td>
<td>37</td>
<td>73</td>
<td>63</td>
<td>73</td>
</tr>
<tr>
<td>Poor</td>
<td>24</td>
<td>60</td>
<td>32</td>
<td>37</td>
</tr>
</tbody>
</table>

1 Bare land value from eqs. (3) and (5) adjusted for site quality, access, and topography. Value supported by DOR.

In structural models such as eqs. (3), (5) or (7), the interpretation of individual regression coefficients is of prime importance. Not only does this guide the model specification exercise, it also requires that care be taken to ensure that individual regression coefficients are as precise as possible. The adverse effect of multicollinearity is the production of regression coefficients with very large standard errors. As shown in Tables 1, 2, and 5, two statistics are used when investigating the degree of multicollinearity present in the 1977 western Washington land sales data: (1) variance inflation factors (VIF), and (2) the determinant of the correlation matrix. VIF's are the diagonal elements of the inverse of the correlation matrix, and are directly tied to the standard error of the regression coefficients. Marquardt (1970) proposes that maximum VIF's should range from one to ten, and closer to one if possible. If perfect orthogonality exists, all VIF's equal one, whereas under collinear conditions, one or more of the VIF's tend toward infinity. For the 1977 western Washington regression analyses, the maximum VIF calculated is 2.513 (left side of Table 5). Other VIF's range from 1.046 - 1.237. Thus, this measure does not indicate that collinearity is a major problem.

In the 1977 court proceedings, the DOR chose to take a weighted average of the bare land values produced by abstraction and regression as shown above. Eq. (7) only received a cursory review. In its ruling, the court stated that the DOR was wrong in using a weighted average produced by two appraisal methods and remanded the resolution of setting bare land values back to the Department. To resolve the dilemma, the DOR ultimately decided to rely solely on the abstraction estimate. However, because the court did not rule on the validity or correctness of either the abstraction or regression methods, the door was left open for further court challenges. Subsequently, a court challenge over the 1978 bare land values was initiated. Seeing an endless succession of such challenges, the DOR sought legislative relief by having the state legislature adopt statutory bare land values in 1981. This schedule, updated annually according to a simple formula, has been in effect for almost nine years and appears to be working satisfactorily. Thus, the argument over the appropriateness of regression as a land valuation tool has subsided.

MULTICOLLINERITY

In structural models such as eqs. (3), (5) or (7), the interpretation of individual regression coefficients is of prime importance. Not only does this guide the model specification exercise, it also requires that care be taken to ensure that individual regression coefficients are as precise as possible. The adverse effect of multicollinearity is the production of regression coefficients with very large standard errors. As shown in Tables 1, 2, and 5, two statistics are used when investigating the degree of multicollinearity present in the 1977 western Washington land sales data: (1) variance inflation factors (VIF), and (2) the determinant of the correlation matrix. VIF's are the diagonal elements of the inverse of the correlation matrix, and are directly tied to the standard error of the regression coefficients. Marquardt (1970) proposes that maximum VIF's should range from one to ten, and closer to one if possible. If perfect orthogonality exists, all VIF's equal one, whereas under collinear conditions, one or more of the VIF's tend toward infinity. For the 1977 western Washington regression analyses, the maximum VIF calculated is 2.513 (left side of Table 5). Other VIF's range from 1.046 - 1.237. Thus, this measure does not indicate that collinearity is a major problem.

The second measure of multicollinearity shown in Tables 1, 2, and 5 is the determinant of the correlation matrix. This measure is close to zero (one) when collinearity is high (low). When the independent variables are normally distributed, Bartlett's (1950) Chi-square test can be used to test the null hypothesis that the determinant is zero (Haitovsky, 1969). In Table 1, the large value of the determinant (.92522) coupled with highly significant regression coefficients indicates that collinearity is not a problem. In Table 2, the determinant is larger (1.78512), but is significantly different from zero. However, the regression coefficient ($\beta_1$) associated with HR, is not significantly different from zero at the .05 level. However, this variable is retained in the model. In Table 5, where two functional forms relating bare land value to site quality, access, and topography are shown, there is apparent collinearity associated with the functional form shown in the right side. For the functional form in the left side of the table, the situation is not so clear. Using Bartlett's test, however, we reject the null hypothesis and conclude that the
determinant \((.38622)\) is significantly different from zero. And, as for the model shown in Table 2, we retain variable \(HR\), even though its regression coefficient is not significantly different from zero.

CONCLUSIONS AND SUMMARY

This paper proposes a regression model for estimating bare land values using forest land sale transactions evidence. Given the constraints of the data collected for each sale, the proposed model (eq. 1) does not contain an intercept term and is designed to allocate the total net sale price across the five value elements associated with each sale. Additional terms reflecting site quality, access, and topography can be accounted for as value modifiers, but it is not appropriate to add them to eq. 1. Lastly, model specifications which force preconceived relationships between the value elements of eq. (1) are rejected as being overly constraining and not reflective of the forest land market.

Because the proposed model [eq. (1)] contains no intercept, and exhibits unequal variation as sale size increases, weighted least squares regression is required. An equivalent transformed model is shown in eqs. (3) and (5). These mixture models allow the use of ordinary least squares. However, the calculation and interpretation of the normal R\(^2\) goodness of fit statistic requires special care in this situation. Alternate means of adjusting R\(^2\) are presented and discussed. In the transformed model [eq. (3) or (5)], all regression coefficients [except \(\beta_0\) in eq. (5)] and the F statistics are significant at the .05 level. And, collinearity is not a problem even though a theoretical restriction in the model suggests a potential problem. This points out that the major portion of the variability in net sale price can be accounted for by the bare land value itself, and not the accompanying immature timber values. It does not imply that the basic model [eq. (1)] is invalid.

To correctly portray the goodness of fit of eq. (1), a pseudo R\(^2\) statistic using the original metric is shown. This calculation demonstrates that the model is doing a good job of predicting net sale price and, more importantly, the bare land value \(B_0\).

Lastly, eq. (7) is proposed to illustrate how site quality, access, and topography can be incorporated directly into the regression analysis. Two functional forms are examined for quantifying the relationship of value across different site quality and access-topography classes. The results using both forms are similar although they differ most on the poor site lands.

Although the DOR no longer uses market transactions evidence to establish bare land values in Washington State, this paper demonstrates the feasibility of such an approach. Further, the use of regression, interjects a degree of objectivity that the previous method (i.e. abstraction) lacked.

LITERATURE CITED


