УДК 338.57:630*67

ANALYSIS OF FACTORS INFLUENCING STUMPAGE PRICES

ГОЙЧУК Д.А., ЗДОБУВАЧ*,
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

We developed a model that determines a variety of factors influencing transaction price of red pine stumpage in Minnesota. Some o the methods used include least squares linear regression, tests on multicolinearity and Goldfeld-Quandt. The model demonstrated that size class, permit duration, and access season have a significant impact on stumpage prices.

Розроблено економетричну модель визначення ціни на соснову деревину від ряду ринкових, адміністративних, та біологічних факторів в штаті Міннесота, США. Використовано лінійний метод найменших квадратів та проведено ряд тестів, зокрема на мультиколінеарність, та Гольдфельд-Квандта. Визначено, що на ціну найбільший вплив мають товарно-сортиментна структура деревостану, тривалість виданого дозволу на рубку, та сезон рубки.

Objective and postulation. The objective of this study is to develop and test a statistical model that can be used to predict stumpage price of pine species. The objective is accomplished through statistical and spatial analysis of data on timber action sales of red pine stumpage on public lands in Minnesota for the period from June 1997 to March 2004 (n=387). The price and geospatial data were obtained from the Minnesota's Department of Natural Resources.

The assumption is made that the actual value of a specific red pine stand may be influenced by the following factors: stumpage category (pulpwood, saw timber or bolts), permit duration, sale volume of forest, distance to the nearest mill buyer, seasonal operability and legal limitations on harvesting operations.

Literature review. One of the first attempts to apply multiple regression analysis as a methodological tool for timber price modeling was made by Anderson's (1969). His research involved transaction data on private bids in South Carolina for the period 1948-1967. Although fifteen independent variables were initially included in the analysis, only eight were tested to be significant, such as lumber price, lumber price change, location,

---

* *Науковий керівник – Кваша С.М., д.е.н., професор, член-кор. НААН України*

time element, and several volume parameters. The coefficient of determination for the model was about 44 percent [1].

More recently, Burak (1996) created an OLS multiple regression model used in the appraisal of pine sawtimber stumpage in the Southern United States. The model utilized private timber sale data from Florida for the period 1985-1990. Six characteristics were determined to be significant estimators, including sawtimber tree size, sawtimber grade of quality, pine pulpwood volume per acre, total hardwood volume, surface and topographic logging conditions, $R2 = 49\%$ [3].

Bare and Smith (1999) attempted to estimate the stumpage value of individual species and qualities in lump-sum timber sales based on western Washington sales data. Ordinary least squares regression included the following variables: hauling distance, logging conditions, volume per acre, four timber quality classes, and time variables. Their corrected for the mean R2 was 71.5% with $F = 56.53$ ($p=0.000$). Nagubadi and Munn (1999) examined the hardwood stumpage market in the South Central United States for the period of 1981-1996. Three stage least squares regression was used to determine parameters of pulpwood and sawtimber stumpage markets for hardwoods [2].

There have been a few documented attempts to apply regression methods for red pine and other species stumpage trend estimation in Minnesota. Merzenich (1986) used multiple regression equations to predict stumpage price based on the characteristics of past sales [5]. MacKay and Baughman (1996) conducted similar research developing a transactions evidence appraisal system for timber tracts administered by the Minnesota Department of Natural Resources. Using thirteen variables, their model explained approximately 83% of the variation in price, with $F= 258$ [4].

Empirical specification of the model and data. The regression model to predict stumpage price based on past transactions may be written as:

$$
\begin{cases}
P_{RP} = f\,(B_{RP}, A_{RP}, M_{RP}) \\
B_{RP} \in [AREA, V, S] \\
A_{RP} \in [PERM, O] \\
M_{RP} \in [D]
\end{cases}
\qquad (1)
$$

where B_{RP} are for biophysical characteristics of red pine stand;

A_{RP} – administrative parameters;

M_{RP} - market-related parameters.

Biophysical B_{RP} price determinants include total area of sale AREA, total volume sold V, and size-class of red pine trees S, represented in the form of product specification (pulpwood, saw timber, wood run). Both total area and volume of sale reflect economies of scale, and therefore, the expected sign of the slope coefficient for both parameters is positive. Buyers are more likely to be more interested in one large-scale harvesting rather than in several small-scale operations due, due to reduced equipment transportation costs. The expected sign for saw timber and wood run is positive, because they had broader appliance then pulpwood, and more likely to be converted into high value solid wood products.

Administrative A_{RP} variables include restrictions on seasonal operability O, and permit duration PERM. The sign of the slope coefficient is more likely to be positive in cases where operations are not restricted to frozen ground conditions, which allow more flexible technological scheduling. Longer permit durations determine more preferable settings for harvest planning and are expected to have a positive sign. The M_{RP} parameter is the distance to buyer. Its slope is expected to be negative. That is, the smaller the distance, the larger is price. We considered 10 explanatory independent variables, six of which are dichotomous variables (Table 1).

Analysis and results. Multiple OLS regression models were used to study relationships between explanatory variables and stumpage price. Dichotomous variables were introduced to represent seasonal operability and stumpage product subgroups in the regression model. The permit duration parameter was derived as a difference between sale date and permit expiration date. Volume and price characteristics for all product categories were converted into single units of measure (i.e., cords; $ per cord) using conventional cord conversion factors. Red pine prices were deflated to common scale using Consumer Price Index.

Table 1.

Codification and characteristics of variables

Code	Description	Units	Data Summary			
			Mean	**Std. Dev.**	**Min**	**Max**
P	Stumpage price	$/cord	48.94	38.38	2.5	240
AREA	Total sale size	acres	413.14	352.94	1	2730
SAW	Sawtimber (YES if SAW=1)	binary	…	…	0	1
PULP	Pulpwood (YES if PULP=1)	binary	…	…	0	1
WR	Wood Run (YES if WR=1)	binary	…	…	0	1
PERM	Permit duration	years	4.01	1.24	1	6
W	Winter access (YES if W=1)	binary	…	…	0	1
SP	Spring access (YES if SP=1)	binary	…	…	0	1
SU	Summer access (YES if SU=1)	binary	…	…	0	1
V	Total volume sold	cords	1472.2	2457.4	1	25900
DIST	Distance	km	62.46	55.21	1.5	318

Several hypotheses were tested in this study. First, it was interesting to test whether each of the identified variables were significant, and of the expected sign. Second, three product (SAW, PULP, PW) and three season (W, SP, SU) dummy variables were separately tested whether they are jointly significant, and have similar slope coefficients. Third, test of heteroscedasticity for suspected variables completed using Goldfeld-Quandt test and corresponding F-statistic. Forth, the model was checked for multicollinearity. Finally, it was tested whether the initial model with 11 variables is not significantly better than the final model with reduced number of variables. The above hypotheses are summarized in the classical form in Table 2.

Table 2.

Study Hypotheses

Null Hypothesis	Restrictions	H
LNAREA, LNVOLUME, SAW, WR, PERM, SU. Separately each slope coefficient is positive.	H_0: $\beta_i > 0$	H_0 is rejected
LNPDIST, W, SP, PULP. Separately each slope coefficient is negative.	H_0: $\beta_i < 0$	H_0 is accepted
Constant, LNAREA, LNVOLUME, LNPDIST, W, SP, SU, SAW, WR, PULP, PERM. Separately each coefficient is equal to zero.	H_0: $\beta_i = 0$	H_0 is rejected
W, SP, SU. Jointly equal to zero.	H_0: $\beta_w = \beta_{sp} = \beta_{su} = 0$	
SAW, WR, PULP. Jointly equal to zero.	H_0: $\beta_{saw} = \beta_{wr} = \beta_{pulp} = 0$	
Multicollinearity	H_0: $e_1, e_2, \dots, e_n$ independent	H_0 is accepted
Heteroscedasticity	H_0: Var(ei) = Var(ej).	
Reasonable for practical purposes R^2	$R^2 \geq .55$	

Individual histograms of P, AREA, DIST, and VOL did not illustrate normal distribution, and they have positive skew. A switch to natural logarithmic scale was made with an introduction of corresponding variables LNP, LNAREA, LNDIST, and LNVOL. The data was checked for linearity and normality; identification and exclusion of outliers was done using box-plot (Figure 1). Total number of outliers removed n=12 out of 387. The new total of observations was 375.

The bivariate correlation was conducted between dependent and all explanatory variables. Pearson correlation and two tailed test of significance were used for continuous variables that are assumed to follow a Gaussian distribution (LNDIST, LNAREA, LNVOL). The nonparametric Spearman correlation test was applied to other variables with no violation of explanatory variable independence was detected. Correlation is significant for PULP, SAW, LNVOL at the 0.01 level of significance, for PERM, WR, W at the 0.05 level. The relationship is positive for SAW, WR, LNVOL and PERM and negative for other variables, as expected.

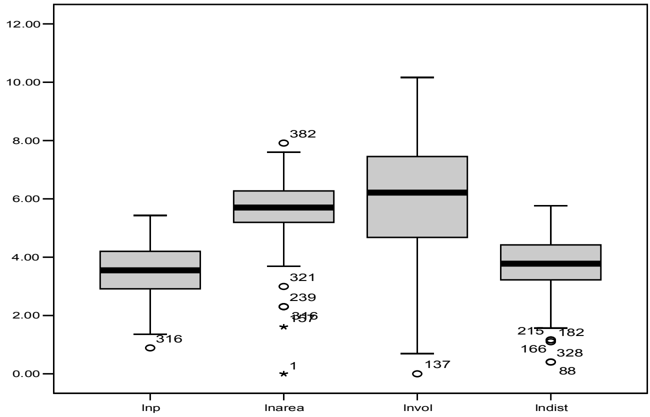

Figure 1. Box-plots of LNP, LNAREA, LNDIST, and LNVOL

Visual examinations of scatter plot of standardized residuals vs. predicted value and scatter plot of standardized residuals vs. each continuous independent variable do not indicate non-linearity or heteroscedasticity of the model (Figure 2). To verify visual observation for LNVOL, Goldfeld-Quandt test was run. The observations were ordered in ascending order of suspected explanatory variable LNVOL and p central observations were omitted (p = n/3 = 375/3 = 125) to catch differences in variances. The separate set of regressions was fit to both set of observations and the test F statistic was calculated. Goldfeld-Quandt test did not detect heteroscedasticity at the 0.01 level of significance, with F=23.223, indicating no need for the White estimators.

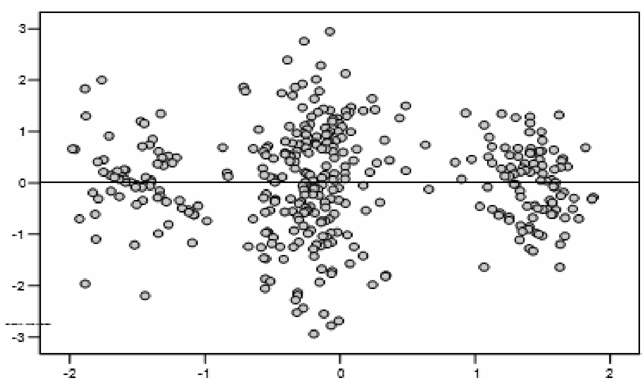

Figure 2. Heteroscedasticity test results

The regression was run for LNP, LNAREA, LNVOL, LN and explanatory variables from Table 1. Model was created using approximately 80% of randomly selected transactions (308 out of 375). The estimated coefficient of determination R^2 = .609, adjusted R^2= .596. The standard error of the estimate was equal to 0.50528. Analysis of VIF, condition index and tolerance statistic for each parameter showed that multicollinearity was not an issue.

The reduced regression model (2, 3) contained five explanatory variables: SAW, PULP, WR, PERM, and W. The estimated coefficient of determination R^2 = .599, adjusted R^2=.592. The standard error of the estimate was equal to .50757. The coefficients are shown in Table 3. Each variable was statistically significant at 0.01 level. F-statistic for the model is 90.403, which was statistically significant at the 0.01 level of significance.

$$LN(P) = 3.331 - 0.740 * PULP + 1.011 * SAW + +0.615 * WR + 0.071 * PERM - 0.269 * W \quad (2)$$

$$P = 3.331 * PULP^{-0.740} * SAW^{1.011} * WR^{0.615} * PERM^{0.071} * W^{-0.269}, \quad (3)$$

where P - selling price, PULP – pulpwood, SAW – sawtimber, WR – wood run, W – winter, and PERM - permit duration.

Table 3

Stumpage price model

	Unstandardized Coefficients		**T**	**Sig**	**Collinearity Statistics**	
	B	**Std. Error**			**Tolerance**	**VIF**
(Const)	3.331	.125	26.720	.000		
pulp	-.740	.079	-9.387	.000	.881	1.135
saw	1.011	.070	14.443	.000	.894	1.118
wr	.615	.213	2.888	.004	.965	1.036
perm	.071	.024	2.915	.004	.952	1.050
w	-.269	.076	-3.518	.001	.974	1.027

The F-test was also used to specify whether the model with ten independent variables was better than optimized model with five independent variables. Applied F-test restrictions: $\beta_{LNAREA} = 0$, $\beta_{SP} = 0$, β

$_{SU} = 0$, $\beta_{LNVOL} = 0$, $\beta_{LNDIST} = 0$ The p-value is larger than the chosen significance level, therefore at significance level 0.05 < 0.8670, thus hat there is no basis for choosing the general rather than the optimized model.

The reduced model was created using approximately 80% of randomly selected transactions (308 out of 375). Its reliability was tested on the remaining 20% of cases. An average error of prediction was $3.11 per transaction. F-tests for slope hypotheses were run. The estimated F* was greater than critical F_c value, and p-value was smaller than the chosen significance level 0.05, for tests $\beta_{lnvol}=0$, $\beta_{pulp}=0$, $\beta_{saw}=0$, $\beta_{wr}=0$, $\beta_{perm}=0$, $\beta_{w}=0$ (Table 4). Therefore, there is sufficient statistical evidence to reject H_0: $\beta_i=0$ for these variables (Table 1), showing their importance in the model. At significance level of 0.05 we also conclude that H_0: $\beta_{saw}= \beta_{wr}= \beta_{pulp}= 0$ can be rejected, indicating SAW, WR, PULP dummy variables are important in explaining red pine stumpage prices. H_0: $\beta_{w}= \beta_{sp}= \beta_{su}= 0$ cannot be rejected.

Conclusions. As a result of the study, a linear multiple regression model for red pine transaction price on Minnesota's public land was developed. The model allows calculation of price knowing the values of five variables (sawtimber, pulpwood, wood run, permit duration and winter). At the 0.01 level of significance, the additive combination of the three independent variables jointly accounts for about 60% of the variation in price. The remaining 40 percent of the variation could be explained by additional variables, not tested in my analysis, such as timber quality, market demand, type of logging equipment etc. The estimated elasticity was negative for pulpwood and winter, as expected. The model has limitations due to extrapolation, which occurs when some of the independent variables used for model design are out of range used for model design.

Literature.

1. Anderson, W.C. 1969. Pine Sawtimber Price Behavior in South Carolina / W.C. Anderson. – Research paper SO-129. – New Orleans: USDA Forest Service, Southern For. Exp. Station. – 12 p.

2. Bare, B.B. Estimating Stumpage Values from Transaction Evidence using Multiple Regression / B.B Bare, R.L. Smith // Journal of Forestry. – 1999. – № 97 (7). – P.32-39.

3. Burak, S.G. Pine Sawtimber Stumpage Price Determinants: an analysis of private timber sales in Florida / S.G. Burak, J.P. Caulfield, S.H Bullard, eds. // A world of forestry: proceedings of the 25th annual Southern Forest Economics Workshop, New Orleans, LA. – 1995.

4. MacKay, D.G. Multiple Regression-based Transactions Evidence Timber Appraisal for Minnesota's State Forests / D.G. MacKay, M.J. Baughman //

Northern Journal of Applied Forestry – 1996. – №13(3). – P.129-134.
5. Merzenich, J.P. Predicting stumpage prices and appraising timber using multiple regression equations / J.P. Merzenich [P.E. Dress, Field, R.C., eds.] // The 1985 Symposium on Systems Analysis in Forest Resources: Proceedings. – 1987. – P. 508-519.
6. Nagubadi, V. An econometric study of the hardwood stumpage market in the South Central United States: preliminary results / V. Nagubadi, I. Munn // Proceedings 1998 Southern Forest Economics Workshop. – 1998. – P. 185-190.