A regression analysis exploring the impact that sawmills and production have on Southeast softwood timber prices

By Christopher Humphrey
6/24/2015

Submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the requirements for the degree of
Master of Natural Resources
Economics and Management Technical Option

Bob Abt, Chair
Dennis Hazel
Stacy Nelson

Abstract: This analysis focused on the price and production of softwood saw timber in the states of North Carolina, South Carolina and Virginia. Since the preponderance of softwood timber production occurs along the eastern portion of these states, only specific regions were used for modeling purposes. The data was analyzed via a 2 stage least squares regression model which determined the relationship and effect between the number of sawmills, and estimated production within in a particular region and the price for timber. After analyzing the data, it was determined that the positively correlated relationship that exists amongst the independent variables in the time series format does not exist when a regional, spatial component is taken into account. The negative relationship exists between prices and markets on a regional level at the 98% confidence interval. The elasticity of sawmills was determined to be a 1% increase in sawmills leads to a 5% decrease in prices.

Introduction

Timber markets in the southeastern US have seen a drop in prices over the years since 2001. During this same period, many sawmills have closed and timber production has also fallen except for the recent upward trend since 2010. Analyzing these market dynamics can give individual landowners a better understanding of the value their timber assets. This project will examine the relationship between the number of sawmills (markets), and the prices and production of softwood Saw timber in specific regions within the states of Viginia, North Carolina, and South Carolina. This relationship will be determined through a linear regression analysis where softwood sawtimber prices are the dependent variable and sawmills and softwood sawtimber production are the independent variables. The direction of the coeffcients for sawmills and production will determine the relationship to prices and a log-log model of the linear regression to determine the elasticity of each of the respective independent variables.

<u>Modeling Regions</u>

The data utilized for this report covers the states of Virginia, North Carolina, and South Carolina. The regression analysis is delineated along the price regions defined by Forest 2 Market, as displayed in Figure 1. The sawmill information as well as the production data is supplied by Southern Research Station (SRS) Timber Production Output (TPO). The analysis begins at the county level for all states (246 in total), and is then analyzed at the Forest 2 Market regional level. The years covered in the report are from 2001 to 2011. However, because of the nature of the TPO data, only the data from odd years could be used. The total number of regions modeled that were six yielding for total of 36 observations.

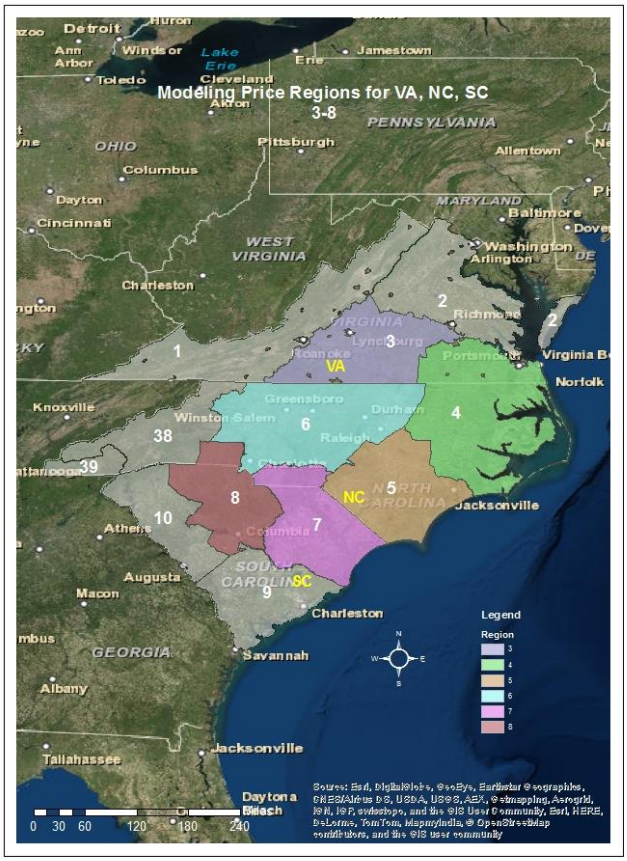


Figure 1: F2M Regions used for price model

Price Data

The data obtained from Forest 2 Market is an average of semi-monthly prices of pine saw timber per ton of the given region each odd year. Figure 2, displays a map of the model price regions and average prices of saw timber for the odd years 2001-2011. For the given model regions, prices are lower on average in central Virginia (region 3) and central North Carolina (region 6) and higher on average in South Carolina (Regions 7 & 8).

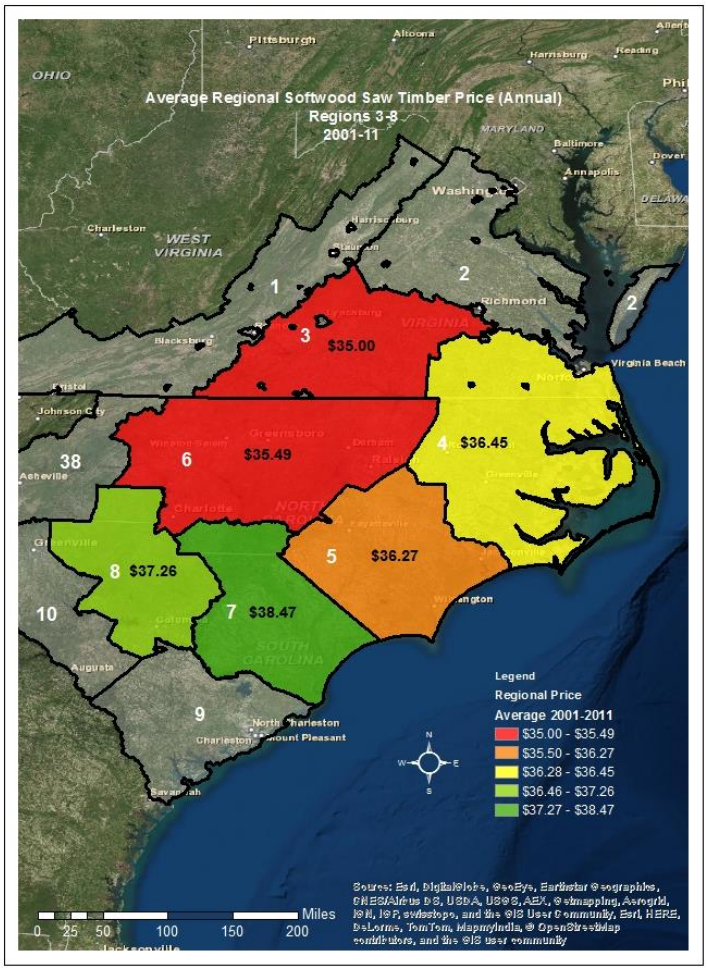


Figure 2: Average Regional Price (2001-2011 odd years) for Regions 3-8

Production data

The TPO production data consists of county production of softwood sawtimber within the model areas. The data shows a trend of higher softwood production along the Atlantic coast versus the western portion of modeled regions, with the exception of region 6. These regions have averaged the majority of production for the given years within the dataset.

Figure 4 shows counties with the highest production on average were Beufort and Craven (NC) in region 4 and Georgetown (SC) in region 7. The trend in the data for higher production can be seen along the coastal areas of North and South Carolina.

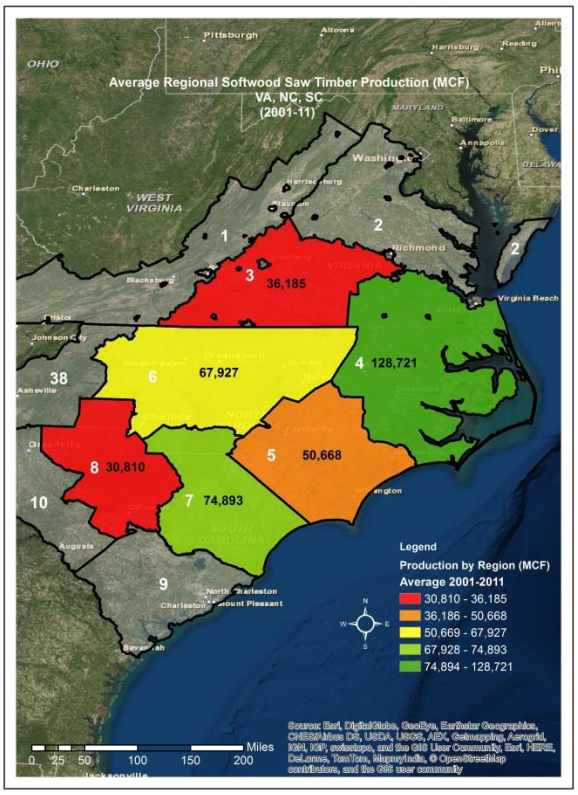


Figure 3: Average Regional Price (2001-2011 odd years) for Regions 3-8

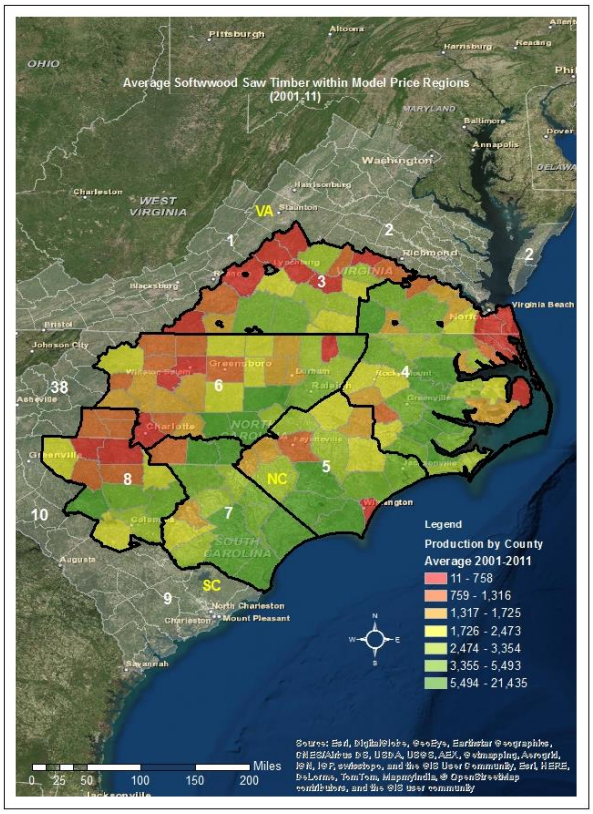


Figure 4: Average softwood sawtimber by county (MCF)

Sawmills

The sawmill data was provided by the Southern Research station. The data consists of a total count of the sawmill survey respondents by county in the states of Virginia North Carolina, and South Carolina. Similar to the production data, the sawmill data is only for the odd years from 2001-2011. It should be noted that this number represents both hardwood and softwood sawmills.

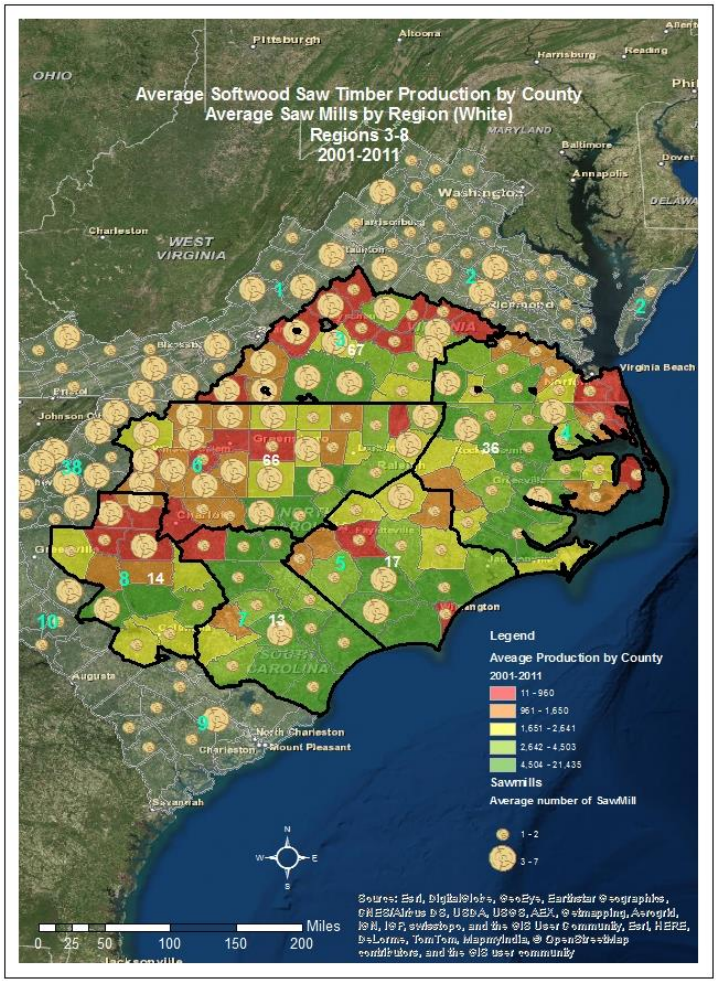


Figure 5: Average Sawmills by Region

The average number of sawmills are concentrated in the western part of the model region where softwood production is lower. Figure 5, displays the average number of sawmills per county overlay with the average timber production by county. In total, Region 3 and 6 have the highest number of sawmills at 66 and 67 respectively. Regions 5, 7, 8 have the fewest on average at 17, 14, and 13.

Overall Market Data

The correlation matrix of the entire model region,Table 1,indicates that the annual price, total sawmills and production positively correlated and the yearly time trend is overhwelmingly negative for given years. The graph in Figure 6 of the price and sawmill relationship exhibits a correlation coefficeint of R^2=0.84. This is a strong positive relationship however this relationship does not appear to be as strong as the relationship betweeen prices and production as can be seen in figure 7. Although this data set is only for 6 years, the trend for both prices, sawmills and production have declined significantly since 2001.

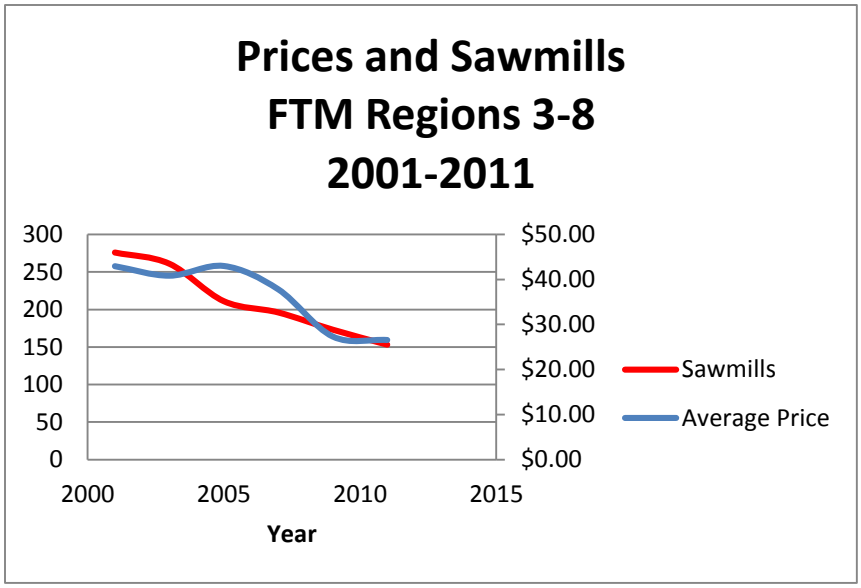


Figure 6: Time series relationship of prices and sawmills

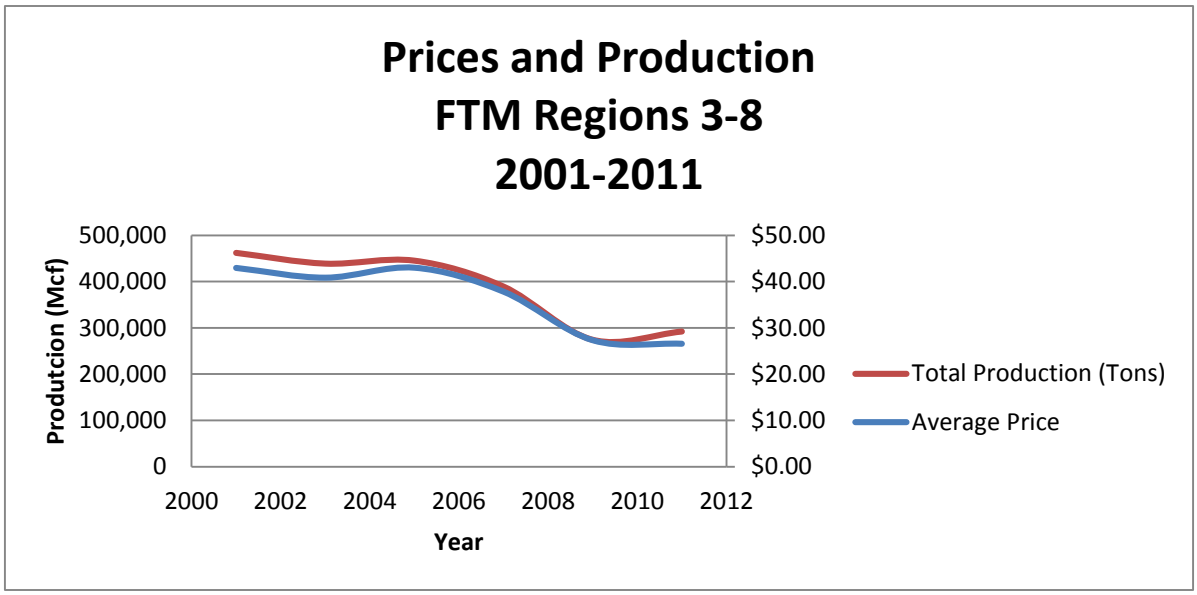


Figure 7: Time series relationship of prices and production

Correlation matrix				
	Year	Average Price	Sawmills	Total Production (Mcfs)
Year	1			
Average Price	-0.900376066	1		
Sawmills	-0.985532668	0.839158519	1	
Total Production (mcf)	-0.915854697	0.990359769	0.870546302	1

Table 1: Correlation Matrix of Time Series variables

Methods / Procedure

Price = $\beta_0 + \beta_1$ (SawMills $_{Buffer}$) + $\beta_2(\widehat{Production})$+ β_3(Time trend)+ u_i

<u>Sawmills $_{Buffer}$</u>

The SawMills$_{Buffer}$ variable is to give a more accurate count of markets within a given price region. For most regions, there are markets in counties which border along the boundaries of price region. These sawmills markets have ability to procure timber from within the price region and are subject to be influenced by those prices. To incorporate this, a buffer around the individual price regions was created in ArcGIS to include these markets in the overall region count. A Buffer creates a new coverage of buffer polygons around specified input coverage features. In ArcGIS, the buffer tool was used to define an area within specified distances of 5,10,15,20 and 25 miles around each region (ArcGIS).

<u>$\widehat{Production}$ (Estimated Production)</u>

<u>2 Stage Least Squares Linear regression model with a single endogenous regressor</u>

Because of the issue of simultaneity when a timber price and timber production transaction occurs, a 2 stage least squares regression analysis was developed. Calculated in two stages, the first stage decomposes X into two components: a problematic component that is correlated to the regression error and a problem-free component that is uncorrelated with the error. The second stage uses the problem free component to estimate **β_1** (Stock421)**.**

Given the linear equation

Price = $\beta_0 + \beta_1$ (Sawmills) + β_2(Production) + β_3(Time trend)+ u_i

*where **Production** is an endogenous variable*

The first stage regression relates Production to the exogenous variables. An OLS model for Production is developed to obtain fitted values for an estimated value ($\widehat{\mathbf{Production}}$).

1st stage) $\widehat{\mathbf{Production}} = \pi_0 + \pi_1(Z_{1\ldots2\ldots3})$

The second stage of the 2 Stage Least Squares Linear regression mode is estimated by OLS, except Production is replaced by its predicted value from the first stage, $\widehat{Production}$ (Stock 432).

2nd stage) Price = $\beta_0 + \beta_1$ (Sawmills) + β_2($\widehat{\mathbf{Production}}$)+ β_3(Time trend)+ u_i

Where $\widehat{\boldsymbol{Production}}$ represents estimated production

<u>Linear Time Trend</u>

A chronological variable of time, the linear time trend is also known as a trend variable. The coefficient of this variable indicates an upward trend if the sign is positive or a downward trend the sign is negative (Gujarati 148). Denoted as t.

Estimated Production Model Selection Process : *Backwards Elimination*

The explanatory variables estimated timber production model was determined through a backwards elimination regression process at the 99% confidence level. The backward elimination technique begins by calculating statistics for a model, including all of the independent variables. Then the variables are deleted from the model one by one until all the variables remaining in the model produce statistics significant at 99%(.01) confidence level. At each step, the variable showing the smallest contribution to the model is deleted(SAS).

• **DeerPermits : The annual number of deer hunting permits in a region** • **AvgWageFor: The annual average weekly wage of foresters in a region** • **AvgEmployFor: the annual average number of those employed in the forestry profession** • **RegGDP: the annual total GDP for a region** • **Pop: the annual total population in a region** • **Housingstarts: annual housing starts** • **PerInc: Average Regional Personal Income** • **Swap10yr: Annual Daily average of 10yr swap rate** • **Sawmills_buffer: Regional total of sawmills including adjacent counties**	• **ConSent: Annual consumer sentiment** • **Tobacco: Annual regional tobacco production** • **AvgWageTruck: The annual average weekly wage of truckers in the region** • **AvgEmployTruck: The annual average number of those in the trucking profession** • **Diesel: Average yearly diesel price** • **Gold: Average yearly Gold prices** • **PCE : Personal Consumption Expenditures** • **IncAssets : Annual Personal Income: receipts on assets** • **LumberCons : Annual US Softwood Exports**

Saw mill Elasticity (Log-Log Model)

$$\ln(\text{Price}) = \beta_0 + \beta_1 \ln(\text{SawMills}_{\text{Buffer}}) + \beta_2 \ln(\widehat{\text{Production}}) + \beta_3 \ln(\text{Time trend}) + u_i$$

This model is used to determine the percentage change that the explanatory variables have on the dependent variable.

In the log-log model, a 1% change in X (SawMills) will be associated with a β_1 % change in Y (Price). In this respect, β_1 is the elasticity of Y (Price) with respect to X (SawMills). This coefficient will be a ratio of the percentage change in Price associated will the percentage change in Sawmills(Stock 270).

$$\beta_1 = \frac{\Delta\text{Price}/\text{Price}}{\Delta\text{SawMills}/\text{Sawmills}} = \frac{100 \text{ x} \left(\frac{\Delta\text{Price}}{\text{Price}}\right)}{100 \text{ x} \left(\frac{\Delta\text{SawMills}}{\text{Sawmills}}\right)} = \frac{Percentage\ change\ in\ Price}{Percentage\ change\ in\ Sawmills}$$

Multicollinearity : Variance inflation Factor

When multicollinearity is present amongst the independent variables, interpretation of the **β's** can be incorrect. The value of the **β's** may have an opposite sign from what is expected (Menhendhall 347). Multicollinearity amongst the independent variables contributes redundant information. The variance inflation factor (VIF) is a measure of multicollinearity of the individual beta parameters.

$E(y)=\beta_0+\beta_1 x_1 \ldots \beta_k x_k$

$$S^2_{\hat{\beta}_i} = \boldsymbol{s^2}\left(\frac{1}{1-R_i^2}\right) \textbf{ or } (\text{VIF})_i = \frac{1}{1-R_i^2}$$

Where s^2 is the estimate of σ^2, the variance of ε, and $R_i^2$is the multiple coefficient of determination for the model that regresses the independent variable on x_i on the remaining independent variables $x_1, x_2 \ldots x_{i-1}, xi+1, \ldots x_k$.The quantity $\frac{1}{1-R_i^2}$ is called the Variance inflation factor (VIF) for the parameter β_i. A VIF >10 indicates multicollinearity is present and may cause problems (Mendenhall 348). If multicollinearity is present, removing the problematic variable will be employed as a strategy to overcome this.

Durbin Watson Test and Lag model

The Durbin-Watson (DW) test is a statistic used to detect auto correlation amongst the error terms of the model. DW values at 2 indicate no autocorrelation. Values near 0 indicate strong positive correlation in error terms.

$$d = \frac{\sum_{t=2}^{n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{n} \hat{\varepsilon}_t^{\,2}}$$

Where the range of d: $0 \leq d \leq 4$

When autocorrelation is present, the t-values of the β statistic can be inflated (Mendenhall 507).

Results

Sawmill Statistic and Buffer

The sawmill variable is a total count of sawmills within a price region. However, since sawmills neighboring the border of price regions can competitively acquire wood from inside of the price region from which they operate, an additional "buffer" zone was extended around the region to include overlapping sawmills within the total count

Buffer distances were measured in intervals of 5, 10, 15, 20, and 25 miles outside of the border of the price regions. A backwards elimination of the number of sawmills in each of the buffer distances was used in SAS to the best distance to use in the price model. The result was a buffer distance of 20 miles surrounding the regions with the exception of region 4 where not using a buffer proved to be more accurate.

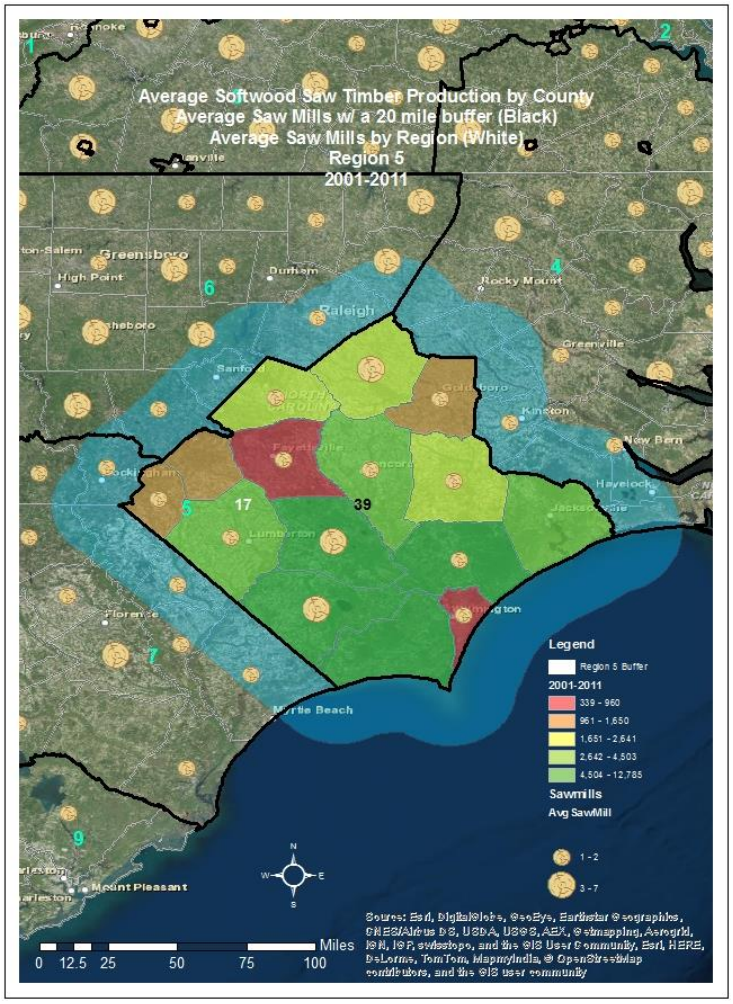


Figure 8: Region 5 with a 20mi buffer

Here in Figure 8, the map displays the number of sawmills by county overlay with the regions average production by county. The blue area represents the extended buffer of 20 miles surrounding the region. The total count of sawmills using to buffer is 39 whereas without the buffer the count is only 17.

Region 4 without a buffer

The location of region 4 makes it unique as it is the region with the largest coastline. As shown in Figure 9, the coastal part of this region begins as far north as Virginia Beach, VA to Morehead City, NC including the outer banks of North Carolina. The distribution for the majority of sawmills in region 4 appear to be located closer to the coast. This implies that the actual number of sawmills affected by regional pricing does not necessarily include counties outside of the border. By solely using the count of sawmills within region 4, the number of markets statistic is a more accurate representation of the data.

Figure 10 shows the final results of the $\text{Sawmill}_{\text{Buffer}}$ statistic. The numbers in white represent the average number of sawmills in the price region without using a buffer. The numbers in black represent the average number of sawmills in the region including a 20 mile buffer.

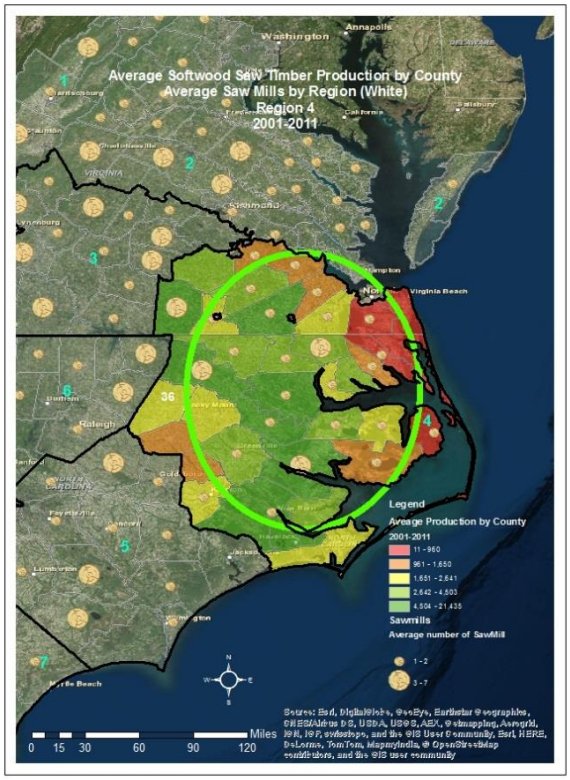


Figure 9: Region 4 sawmill distribution

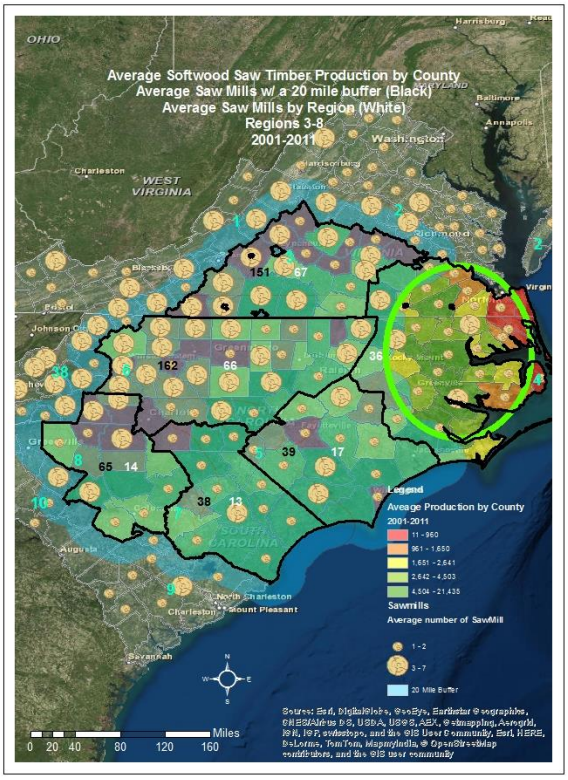


Figure 10: Project region with 20 mi sawmill buffer

Backwards Elimination Process

When the backwards elimination process was ran, the sawmills variable was included as a possible explanatory variable. The process determined that Sawmills were not a good predictor of production (Figure 11). This could imply that sawmill production capacity is not being fulfilled as there is no discernable relationship to production and as such, is not significant in determining the quantity of timber produced in a region. This could be attributed to the nature of the sawmill data as it includes hardwood sawmills and not solely softwood.

Summary of Backward Elimination									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1		PerInc	Annual Regional Per capita Income	13	0.0001	0.7347	13.0050	0.01	0.9443
2		SawMills_buffer	Annual Regional Total of Saw mills including adjacent counties	12	0.0059	0.7287	11.4743	0.49	0.4906
3		AvgEmployFor	Annual Regional Average Employment: Agriculture, Forestry, Fishing & Hunting	11	0.0056	0.7231	9.9179	0.48	0.4976
4		LumberCons	Annual US Softwood Exports	10	0.0208	0.7023	9.5646	1.80	0.1919
5	Gold		Annual Monthly Average Gold Price	11	0.0208	0.7231	9.9179	1.80	0.1919
6		Gold	Annual Monthly Average Gold Price	10	0.0208	0.7023	9.5646	1.80	0.1919
7	HousingStarts		Annual Housing Starts	11	0.0208	0.7231	9.9179	1.80	0.1919
8		HousingStarts	Annual Housing Starts	10	0.0208	0.7023	9.5646	1.80	0.1919
9	DieselPrice		Annual Weekly Average Diesel Price	11	0.0208	0.7231	9.9179	1.80	0.1919
10		Swap10yr	Annual Daily Average 10yr Swap Rate	10	0.0138	0.7094	9.0066	1.19	0.2857
11	ConSent		Annual Index of Consumer Sentiment	11	0.0138	0.7231	9.9179	1.19	0.2857
12		DieselPrice	Annual Weekly Average Diesel Price	10	0.0011	0.7220	8.0047	0.10	0.7605
13		IncAssets	Annual National Personal income: receipts on assets	9	0.0012	0.7208	6.1011	0.11	0.7434
14		CanadaProd	Annual Candaian Timber Production	8	0.0226	0.6982	5.8927	2.11	0.1585
15		ConSent	Annual Index of Consumer Sentiment	7	0.0228	0.6753	5.7011	2.04	0.1643
16	year		Linear time trend	8	0.0231	0.6985	5.8714	2.07	0.1618
17		year	Linear time trend	7	0.0231	0.6753	5.7011	2.07	0.1618
18		Tobacco	Annual Regional Tobacco Prodcution	6	0.0230	0.6524	5.5181	1.98	0.1704
19		AvgWageTruck	Annual Weekly Regional Average: Transportation and Warehousing	5	0.0148	0.6376	4.6896	1.23	0.2756
20		PCE	Annual Durable Goods Consumption	4	0.0613	0.5763	7.5386	5.07	0.0318

Figure 11: Backwards elimination variables removed

Predicted production model with high variance inflation

The first production equation derived from the backwards elimination process is shown in Figure 12. The model had an overall F-Statistic 16.1, and an R^2= 0.729. However, the population variable contains a high variance inflation factor(VIF) of 23.225. This high VIF indicates severe mullticolinearity and would present a problem as an explanatory variable in a predictor model. To avoid issues that multicollinearity can present, the backwards elimination process was ran again without the population variable.

Softwood Timber Production model Data: 2001 - 2011(odd years) regions: 3 4 5 6 7 8

The REG Procedure
Model: MODEL1
Dependent Variable: Production

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	31797646160	6359529232	16.10	<.0001
Error	30	11846844291	394894810		
Corrected Total	35	43644490451			

Root MSE	19872	R-Square	0.7286
Dependent Mean	63952	Adj R-Sq	0.6833
Coeff Var	31.07332		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	212234	36761	5.77	<.0001	0
DeerPermits	Annual Regional Deer Permits Issued	1	2.39256	0.35611	6.72	<.0001	3.96536
AvgWageFor	Annual Weekly Regional Average Forestry Wages	1	-386.68562	73.68908	-5.25	<.0001	1.58657
AvgEmployFor	Annual Regional Average Employment: Agriculture, Forestry, Fishing & Hunting	1	3.77619	1.18768	3.18	0.0034	1.36857
Pop	Annual Regional Population	1	-0.07135	0.01258	-5.67	<.0001	23.22515
RegGDP	Annual Regional GDP	1	0.96607	0.17362	5.56	<.0001	16.51217

Figure 12: Production model containing variables with high VIF (circled in red)

Production model

Model: $\widehat{\text{Production}} = \beta_0 + \beta_1(\text{Deer Permits}) - \beta_2(\text{Weekly Wage of Forestry}) + \beta_3(\text{Employment in transportation}) - \beta_4(\text{Regional GDP}) + u_i$

The second model produced from the backwards elimination process in SAS is shown in Figure 13. The predictive model has an F statistic of 10.54, and an $R^2$of 0.573. The two strongest model variables were Deer Permits and the Average Weekly wage in Agriculture, Forestry, Fishing & Hunting. The Deer permit data was obtained from the Department of Natural resources from each of the respective states in the model region. The wage data, as well as the data from the other variables in the model, were obtained from the Bureau of Labor statistics. The other variables included the average

Softwood Timber Production model Data: 2001 - 2011(odd years) regions: 3 4 5 6 7 8

The REG Procedure
Model: MODEL1
Dependent Variable: Production

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	25154392880	6288598220	10.54	<.0001
Error	31	18490097571	596454760		
Corrected Total	35	43644490451			

Root MSE	24422	R-Square	0.5763
Dependent Mean	63952	Adj R-Sq	0.5217
Coeff Var	38.18878		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	239456	39420	6.07	<.0001	0
DeerPermits	Annual Regional Deer Permits Issued	1	1.61044	0.30077	5.35	<.0001	1.87273
AvgWageFor	Annual Weekly Regional Average Forestry Wages	1	-467.54295	87.52732	-5.34	<.0001	1.48199
AvgEmployTruck	Annual Regional Average Employment: Transportation and Warehousing	1	-1.87953	0.51740	-3.63	0.0010	9.35540
RegGDP	Annual Regional GDP	1	0.52975	0.15279	3.47	0.0016	8.46571

Figure 13: Estimated Production equation

employment of the Transportation and Warehousing by region and the regional GDP

In comparison to the model in Figure 12, the R^2 decreases from 0.7286 to 0.5763 and the F-test drops from 16.10 to 10.15. The employment of forestry workers variable (AvgEmployFor) is replaced with the average employment of truckers (AvgEmployTruck). The multicollinearity amongst the variables is also reduced but is still present in the Regional GDP and the trucking and warehousing variable. However since the VIF for both variables were less than 10, the model was still used.

Softwood Timber Production model Data: 2001 - 2011(odd years) regions: 3 4 5 6 7 8

The REG Procedure
Model: MODEL1
Dependent Variable: Production

Durbin-Watson D	0.953
Pr < DW	<.0001
Pr > DW	1.0000
Number of Observations	36
1st Order Autocorrelation	0.503

Note: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Figure 14: Durbin-Watson results of production model

Although this new model in Figure 13 model has lower overall significance and correlation, when this model is used in the price model, model is improved in overall R^2, F-stats and higher values for individual beta coefficients.

This predicted production model does exhibit some multicollinearity between the average regional employment in Transportation and Warehousing and regional GDP. The production model also exhibits a positive autocorrelation of the error terms with a DW statistic of 0.953 (Pr<DW : <.0001) as shown in Figure 14.

Strongest Model Variables

Deer Permits

The strongest independent variable is the Deer Permits variable with an overall t-value of 5.35. This implies that number of deer permits within a given region has a positive relationship with timber production.

As show by the map in Figure 15, the number of Deer permits is dramatically less in Virginia than in either North Carolina or South Carolina. This may be due to cultural differences or state laws, regardless, the deer permit data is skewed for the state of Virginia. This however does correspond to lower overall softwood timber production in the state as compared to other states.

The data in Figure 16 indicates that there is an urban versus rural correlation with deer permits and timber production on a county level. In this distinction, urban would also include military as well. However this relationship does not appear to hold for southern Virginia and northern South Carolina (regions 3 and 8). The deer permit data strongly correlates to production along the eastern portion of the modeled regions. The areas circled in blue indicate counties in which the Deer permit data corresponds positively to the production data. The colors of the counties will match and appears more vibrant. The areas circled in gray indicate counties in which the deer Permit data negatively corresponds to the production data. The colors of these counties will be opposite from each

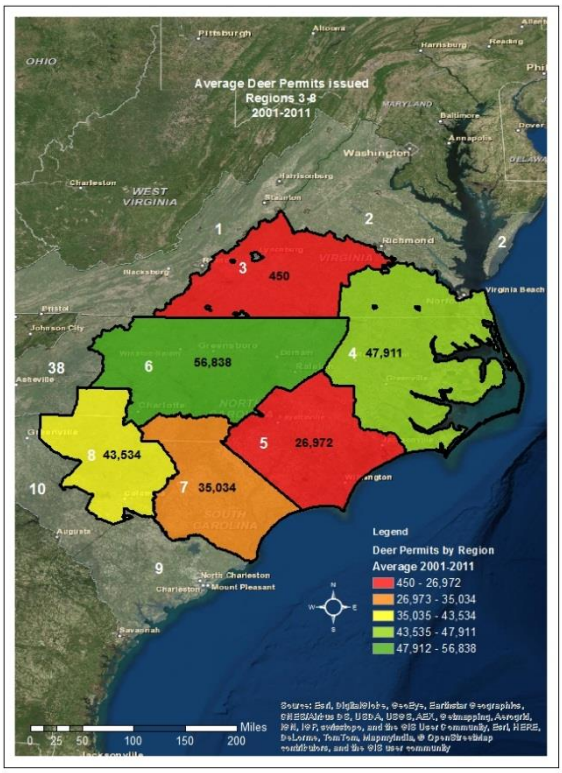


Figure 15: Average Deer Permits by Region

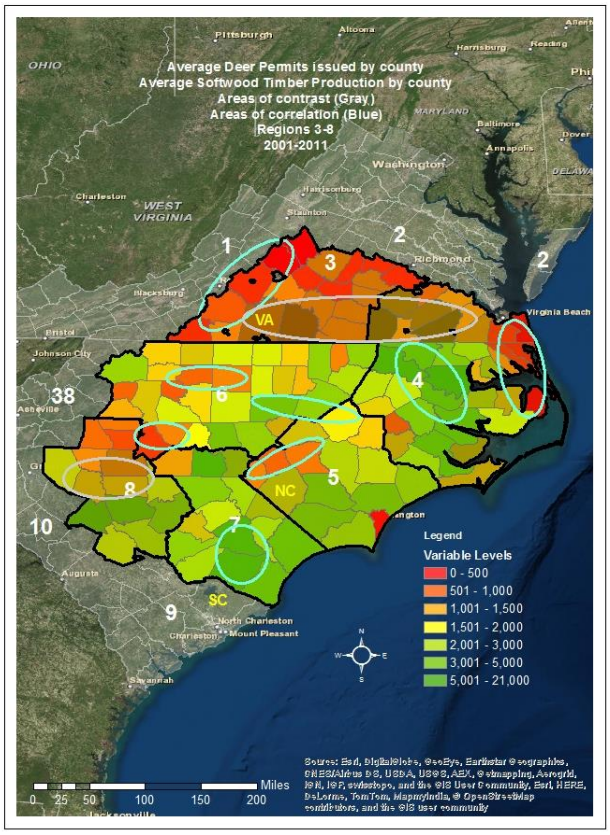


Figure 16: Overlay of deer permits and production by county

other and thus, will contrast with each other appearing brown. The majority of southern Virginia softwood timber production is in stark contrast to the relative issuances of permits.

Weekly wage in Agriculture, Forestry, Fishing & Hunting

The Backwards elimination process determined a negative relationship between the average weekly wage and production. The map in Figure 17 has the color scheme inversed to represent this relationship. The regions with the highest weekly wage were regions 6 and 4. The prices in region 5 were considerably less on average ($441.67) than any other region that was modeled.

The county level wage and production data are represented in Figure 18. The data appears to correlate around urban areas as higher wages are associated with lower production. This is very apparent in region 6 counties Durham, Wake and Mecklenberg. There does appear to be strong areas of correlation in more rural areas for regions 4 and 7 (circled in blue), however overall on average the deer permit data appears to correlate much better.

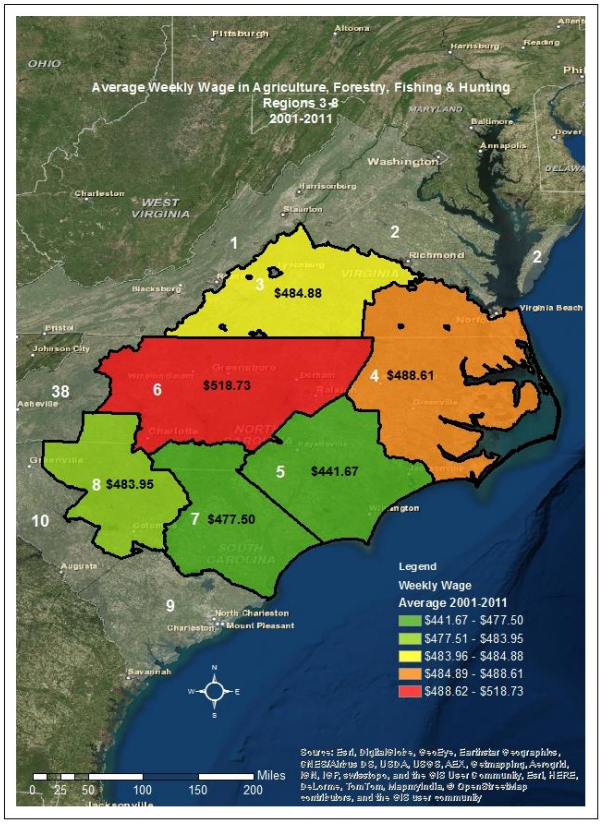


Figure 17: Average Regional Weekly Wage of Forestry, Fishing and Hunting

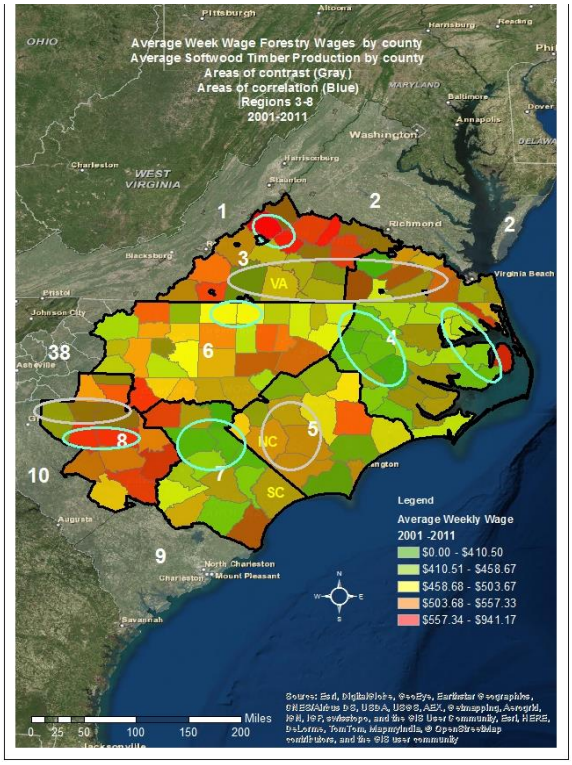


Figure 18: Overlay of weekly wage variable and production by county

Predicted Production vs Actual Production

On average, the production model underestimates Region 3, 4, 7 and overestimates regions 6, 5 and 8. This is more pronounced when looking at the differences in the latest data point, 2011, where the model significantly underestimated region 3's production.

Figure 19 shows the data for the latest year 2011. Here the model extremely underestimates the production in region 3 by 27,000 mcf. The Durbin-Watson statistic from Figure 14 indicated that the model does exhibit positive autocorrelation. The data represented in Figure 18 may exhibit this as the trends in the region seem more pronounced in later data points.

The model on average appears to reasonably predict production for each region as shown in the map overlay in Figure 20. Here both the actual and predicted production data are combined to illustrate areas of contrast and correlation. The more vibrant, correlative regions appear to be Regions 3, 4, and 8, whereas regions 5, 6, 7 appear to be a bit more in contrast.

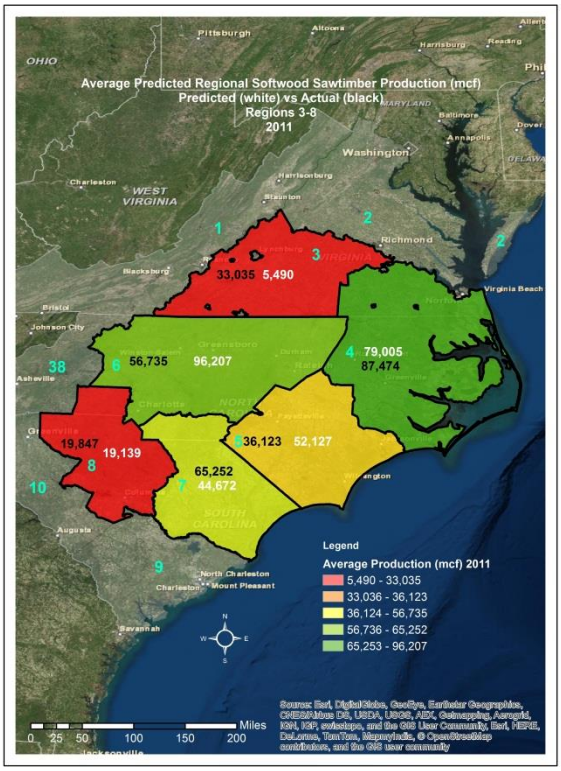


Figure 19: Overlay of predicted production and actual production

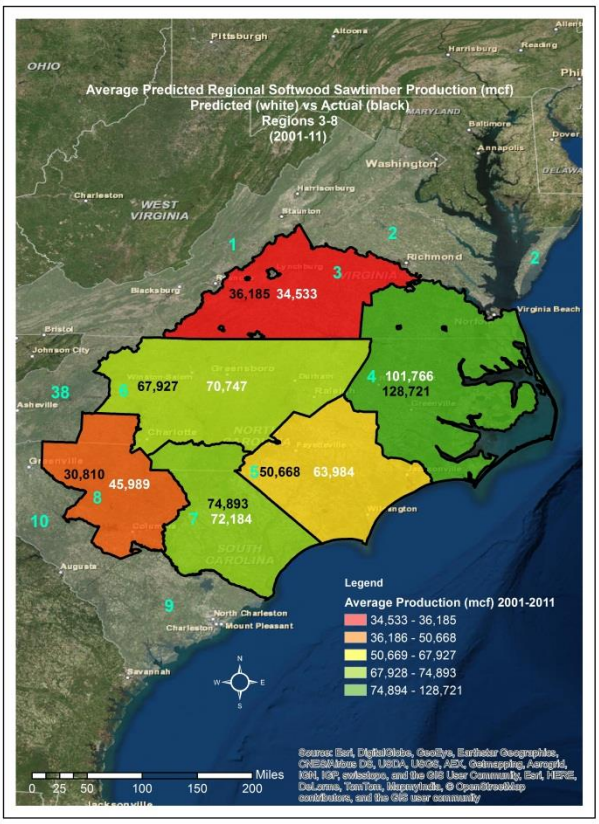


Figure 20: Overlay of average predicted production and average actual production

Price Model

Price = $\beta_0 + \beta_1$ (Sawmill$_{Buffer}$) + β_2($\widehat{\text{Production}}$)+ β_3(Time trend)+u_i

The final result of the entire model building process is shown by the SAS printout in Figure 21. The price model that was derived using the Sawmill$_{Buffer}$ variable, estimated production and a linear time trend contains an overall F statistic of 45.9, an R^2 of 0.81 and a Mean Square Error of $10.77. These statistics indicate that the overall model is significant, the variables are a good fit for the model and the average of predicted price to be within $11.

With a t value of -2.54, the Sawmill$_{Buffer}$ variable has a negative relationship to price at the 98% confidence interval with a low variance inflation of 1.39. The t-value of the predicted production variable (-2.14) also has a negative relationship to price and is significant at the 96% confidence interval.

The linear time trend has the highest t value of either Sawmills or predicted production. This variable also has a negative relationship to price. Thus, the time trend variable can be interpreted as a strong downward trend in prices for the given dataset.

Price-Production (without population as a variable) model Data: 2001-2011 (odd years)regions: 3 4 5 6 7 8

The REG Procedure
Model: MODEL2
Dependent Variable: Price

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1483.06603	494.35534	45.90	<.0001
Error	32	344.64145	10.77005		
Corrected Total	35	1827.70747			

Root MSE	3.28177	R-Square	0.8114
Dependent Mean	36.41917	Adj R-Sq	0.7938
Coeff Var	9.01112		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	4360.60512	404.27838	10.79	<.0001	0
yhat	Predicted Value of Production	1	-0.00005254	0.00002452	-2.14	0.0399	1.52327
year	Linear time trend	1	-2.15279	0.20089	-10.72	<.0001	1.57373
SM20_reg4nB	Annual Regional Total of Saw mills including 20 mile buffer except Region 4	1	-0.02801	0.01103	-2.54	0.0161	1.39592

Figure 21: Price model without the pop variable in production function

Price Model with population included in estimated production

The price model was ran with the estimated production model which contained a high VIF (figure 12). The results shown in figure 22 indicate a price model with a lower F-stat (41.6) and R^2 (0.7959) and higher MSE ($11.66) when compared to the model without the pop variable from figure 21. The model in figure 22 also indicates lower t values for each of the independent variables in the model, the largest difference amongst them is the estimated production variable. With a value of -1.34 (confidence interval at 81%), the production function that contains the population variable is much less significant than the value of -2.14 (confidence interval at 96%) from figure 21. Thus, even though the estimated production function that contains population as a variable is overall more significant and a better fit of the data, due to high multicollinearty being present, it is more effective to predict prices by utilizing the model without the population variable.

Price-Production High VIF (with population as a variable) model Data: 2001-2011 (odd years)regions: 3 4 5 6 7 8

The REG Procedure
Model: MODEL2
Dependent Variable: Price

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1454.70343	484.90114	41.60	<.0001
Error	32	373.00405	11.65638		
Corrected Total	35	1827.70747			

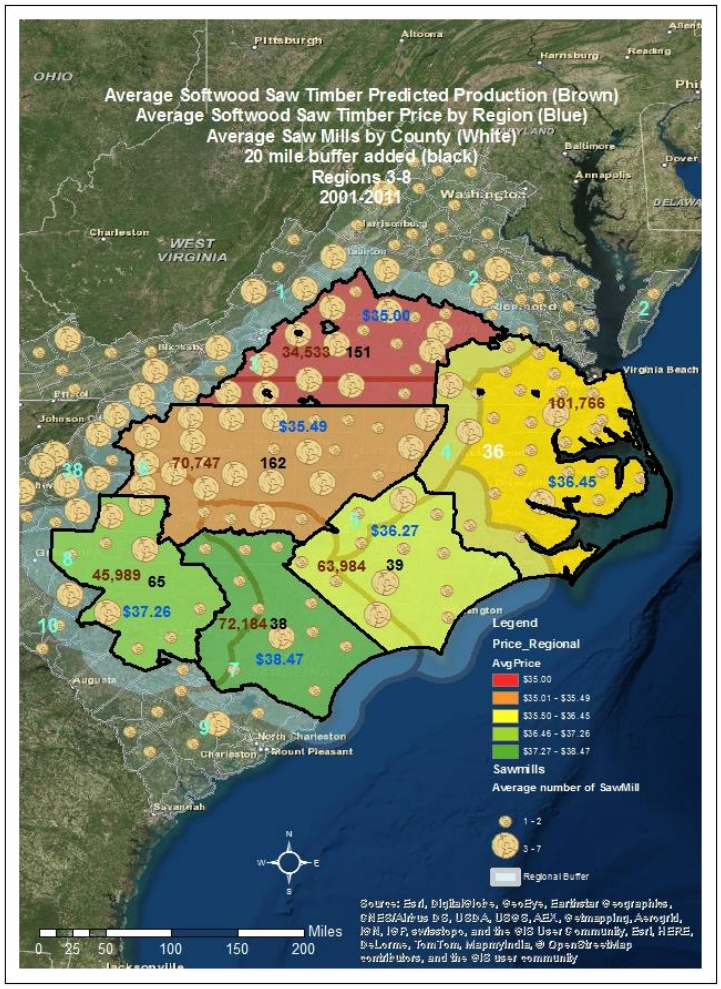
Root MSE	3.41414	R-Square	0.7959
Dependent Mean	36.41917	Adj R-Sq	0.7768
Coeff Var	9.37458		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	4150.50343	403.00103	10.30	<.0001	0
yhat	Predicted Value of Production	1	-0.00002917	0.00002170	-1.34	0.1883	1.40469
year	Linear time trend	1	-2.04898	0.20035	-10.23	<.0001	1.44628
SM20_reg4nB	Annual Regional Total of Saw mills with a 20 mile buffer except Region 4	1	-0.02380	0.01131	-2.10	0.0433	1.35733

Figure 22: Price model with the pop variable in production function

The final result of the relationship between sawmills and prices on a regional level are in stark contrast to the relationship within the overall market. Adding the spatial dynamic to the data by analyzing the regional markets changed the relationship from positive to negative. This may be due to the fact that when the prices area viewed on a regional level, the dynamic between sawmills and prices may be more pronounced whereas when the relationship is viewed from an entire tri-state area, the dynamic may be mitigated when the data is combined and averaged with other market data. For example, the map in figure 23 shows that prices are on average higher in region 7 than in any other region, but also have fewer sawmills. This highlights the negative relationship between prices and sawmills on a regional level but when this data is combined with data from region 3, where the prices are low and number of sawmills are high, the spatial negative relationship would be minimized as the data would cancel out.



The map in figure 23 shows a higher concentration of sawmills toward the western part of the model region where there lower prices as well. Specifically, regions 3 and 6 have the most sawmills yet also they have the lowest prices on average. The predicted production for region 3 is the lowest in the model region as well. However, the negative relationship maybe exaggerated as the Sawmill count will also include hardwood mills increase in number the further west.

Figure 23: Final model map: Price, Predicted Production, Sawmills (20 mi buffer)

Log-Log Model

Log(Price) = $\beta_0 + \beta_1$ Log(Sawmill$_{Buffer}$) +β_2 Log($\widehat{\text{Production}}$)+)+ β_3 Log(Time trend)+u_i

The purpose of the Log-Log Model is to better interpret the beta coefficients. In this formula, the coefficient for Sawmills can be interpreted as the elasticity of price to a change in sawmills. This amount In the Log-Log model, the percentage of price variation that can be accounted for by the number of sawmills is calculated by multiplying the coefficient by 100. In Figure 24, the coefficient of Sawmills is -0.0565 (-5.65%), however, the linear time trend is an overwhelming contributor to the price variation. In this model, the elasticity can be interpreted as a 1% increase in sawmills will result in a decrease in price of 5.65%. The elasticity for predicted production can be interpreted as a 1% increase in predicted production will result in 5.32% decrease in prices.

The REG Procedure
Model: MODEL2
Dependent Variable: logPrice log of price

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.25227	0.41742	39.24	<.0001
Error	32	0.34043	0.01064		
Corrected Total	35	1.59270			

Root MSE	0.10314	R-Square	0.7863
Dependent Mean	3.57401	Adj R-Sq	0.7662
Coeff Var	2.88590		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	939.28916	95.81364	9.80	<.0001
logSawMills	log of Sawmills with a buffer	1	-0.05650	0.03140	-1.80	0.0814
logYhat	log of predicted production	1	-0.05319	0.03316	-1.60	0.1185
logYear	log of linear time trend	1	-122.94895	12.56801	-9.78	<.0001

Figure 24: Log-Log Price model to determine elasticity

Conclusion

This report examines the relationship of the prices of softwood saw timber to the number of sawmills and softwood saw timber production within a given region in the states of North Carolina, South Carolina and Virginia. After analyzing the data, it is determined that the positively correlated relationship that exists amongst the independent variables in the overall time series format does not exist when a regional, spatial component is taken into account. Since only the odd years from 2001 to 2011 were used, much of the change in the variable relationship is possibly due to the size of the data set, as the large differences within data on a regional level are less pronounced when combined and averaged with other regions, changing the relationship of the variables. It is observed that in both the overall time series and regional analysis, the annual negative time trend in prices dominates the significance of all other variables. However, even when the linear time trend is taken into account, the relationship between the number of sawmills and prices for softwood timber is still negative in the southeast region and the effect that a 1% increase in sawmills has on prices is roughly a 5% decrease for the given dataset.

The types of sawmills counted in this report consisted of both hardwood and softwood mills whereas the prices and production data are strictly softwood. When determining an estimation model for production, it is observed that the sawmills variable is not a good predictor. The inclusion of the hardwood mills may distort the values of this variable and contribute to slack present in sawmill utilization. Determining the impact of including the hardwood mills had on the variable is difficult as there is more hardwood production located in the western part of the model region where softwood production is low. In this case, the spatial component may have actually contributed to distortions in the variable statistic as the count of the number of sawmills in the western part of the model regions is inflated by the hardwood mills. Thus, when modeling softwood timber prices, the count of both hardwood and softwood sawmills within a given region may be irrelevant, as the quantity produced is irrespective of the number of producers.

The use of a spatial dynamic evaluates the relationship of the market variables more accurately if appropriate data are available. Including strong spatially related variables, such as deer permits, to determine production, specifically in eastern North and South Carolina can aide in model building. However, the inclusion of strong spatially correlated variables, such as population, can also present multicollinearity and can lead to a less effective model. When examining the number of Sawmills within a given region, it is more accurate to include sawmills from adjacent counties but the distribution of sawmills must be taken into account else the variable will not be as accurate. Overall, combining a spatial component to time series data can greatly enhance timber market analysis but due to spatial distributions of the hardwood and softwood saw mills and production the requirements of the data analysis are quite significant. If more data are available, the focus on prices and softwood sawmills in only the eastern counties and regions where production is high can provide for a more accurate model.

Literature Citations

Gujarati, Damodar N. Essentials of econometrics. 4th ed. McGraw-Hill

Menhdenhall, William. A second course in statistics: regression analysis/ William Mendenhall, Terry Sinich 6th ed. Pearson Education

Stock,James H. Introduction to Economics 3rd ed. Pearson Education Inc. 2011