ABSTRACT

SINGH, NAVINDERPAL. Econometric Analysis of Timber Harvest Behavior and Potential Availability in North Carolina. (Under the direction of Robert C. Abt.)

There has long been an interest in factors which affect the availability of timber inventory for harvest. The main objective of the thesis was to investigate the important factors that impact the timber supply and their effect on the availability. The study involved econometric analysis of timber supply in North Carolina. Binary logistic regression model was used as a statistical tool in this analysis. Timber harvest was formulated as a categorical variable. The model examined the harvest probabilities of forest inventory analysis plots across the study region. The investigation was based upon various biophysical, economic and location predictors. A theoretical model was formulated based upon the production theory and existing literature. The analysis confirmed the results from past studies that biophysical factors like softwood volume, species etc. have major impact on the harvest behavior. Additional variables used in this study as a proxy for age distribution (large tree proportion) and opportunity cost (housing price etc.) also exhibited significant relationship and hence improved the predictive power of the model. Based upon the estimates from econometric analysis the timber availability thresholds were investigated.
© Copyright 2012 by Navinderpal Singh
All Rights Reserved
Econometric Analysis of Timber Harvest Behavior and Potential Availability in North Carolina

by
Navinderpal Singh

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Master of Science in Natural Resources

Raleigh, North Carolina
2012

APPROVED BY:

__________________________________________
Robert C. Abt
Committee Chair

__________________________________________
Fred Cubbage

__________________________________________
Bronson Bullock
DEDICATION

To my family.
BIOGRAPHY

Navinderpal Singh was born and brought up in state of Punjab, India. Agriculture was the family profession for generations. So he did his undergraduate studies in agricultural economics. He also did a graduate degree (MS) in agricultural economics from University of Arkansas, Fayetteville. After studying economics courses and doing applied economics research during his MS, he got further interested in applied economics. To pursue his interest, he did another MS degree in natural resource economics from North Carolina State University, Raleigh.
ACKNOWLEDGMENTS

I would like to thank my major advisor Dr. Bob Abt for his encouragement and direction for this study to be completed. I thank Dr. Fred Cubbage and Dr. Bronson Bullock for serving on my committee and offering helpful comments and critique along the way. I would also like to thank my wife, parents and kid sister for the unconditional support they provided me through my entire life. I must acknowledge my wife and best friend, Aman, without whose love, encouragement and editing assistance, I would not have finished this thesis.
# TABLE OF CONTENTS

## LIST OF TABLES

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>iii</td>
</tr>
</tbody>
</table>

## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>iv</td>
</tr>
</tbody>
</table>

## CHAPTER 1 INTRODUCTION

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

## CHAPTER 2 REVIEW OF LITERATURE

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Theoretical Models</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Empirical Models</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>Wood Availability</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>Summary</td>
<td>16</td>
</tr>
</tbody>
</table>

## CHAPTER 3 METHODOLOGY

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>18</td>
</tr>
<tr>
<td>3.2</td>
<td>Theoretical Model</td>
<td>18</td>
</tr>
<tr>
<td>3.3</td>
<td>Analytical Model</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>Logistic Regression Variables</td>
<td>28</td>
</tr>
<tr>
<td>3.5</td>
<td>Data</td>
<td>33</td>
</tr>
</tbody>
</table>

## CHAPTER 4 RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>36</td>
</tr>
<tr>
<td>4.2</td>
<td>Bivariate Analysis</td>
<td>36</td>
</tr>
<tr>
<td>4.3</td>
<td>Multivariate Analysis</td>
<td>41</td>
</tr>
<tr>
<td>4.4</td>
<td>Results for Availability</td>
<td>44</td>
</tr>
<tr>
<td>4.5</td>
<td>Discussion</td>
<td>48</td>
</tr>
</tbody>
</table>

## CHAPTER 5 CONCLUSIONS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCLUSIONS</td>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>

## REFERENCES

<table>
<thead>
<tr>
<th>Reference</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>REFERENCES</td>
<td></td>
</tr>
</tbody>
</table>
TABLES AND FIGURES

61
LIST OF TABLES

Table 1. Descriptions and summaries of biophysical variables (Continuous) .................. 61
Table 2. Descriptions and summaries of biophysical and location variables
(Categorical) .............................................................................................................. 62
Table 3. Descriptions and summaries of economic variables ....................................... 63
Table 4. Bivariate logistic regression ........................................................................... 64
Table 5. Multivariate logistic regression (complete set of predictors) ......................... 65
Table 6. Multivariate logistic regression (Final) ......................................................... 66
LIST OF FIGURES

Figure 1. Distribution of FIA plots in North Carolina........................................ 67
Figure 2. Box plot for distribution of harvest intensities..................................... 68
Figure 3. North Carolina Physiographic units................................................... 69
Figure 4. Predicted harvest distribution of plots in North Carolina...................... 70
Figure 5. Actual harvest distribution of planted pine plots in North Carolina......... 71
Figure 6. Sawtimber prices versus harvest probability...................................... 72
Figure 7. Slope versus harvest probability......................................................... 73
Figure 8. Softwood proportion versus harvest probability................................. 74
Figure 9. Softwood volume versus harvest probability...................................... 75
Figure 11. Proportion of large trees versus harvest probability.......................... 76
CHAPTER 1 - INTRODUCTION

Forests in the United States are very diverse in composition and distribution and are located across regions with varying biophysical conditions. The forest inventory and analysis (FIA) program of the USDA Forest Service annually report various inventory estimates about volumes, growing stock, mortality and removals etc. But the raw numbers about the inventory volumes reported by FIA do not capture the full set of factors that constrain potential timber inventory. The potential timber inventory or availability in a region is affected by a range of variables such as biophysical characteristics, ownerships characteristics and various economic factors such as timber product prices, harvesting costs etc. For example steep slopes and wet sites make timber harvesting costly and hence constrained. Economic factors can also impact potential supply e.g. timber harvests and replanting were low during recent years due to the historical low prices in the US South, which skewed the age class structure and reduced the availability of younger trees/pulpwood in the near future. Another economic variable that affects potential timber supply is the conversion of timberland to the development land or urban land because of higher opportunity cost. There is a decline in the acreage of forestland near many urban centers as the land is permanently converted to residential, commercial, and other non-forest uses.

Therefore timber supply is a complex function and careful analysis is required to predict the potential supply/availability. A few studies in the past have tried to estimate the harvest probabilities of the forest stands at plot level for example, Polyakov et al. (2010) and Prestemon and Wear (2000). These studies explicitly linked aggregate timber supply models to observations of individual harvest behavior by using FIA expansion factors. The predictors
used in these studies were various site characteristics (e.g. type of soil), location (e.g. physiographic region), and economic variables (e.g. prices).

**Description and Motivation for the Research**

The chief economic product of forests is timber, but the economic benefits, in terms of climate control, pollution abatement, and wildlife maintenance are significant. Timber is used for products ranging from veneer to biomass to mulch.

Forests are also a source of economic, social and environmental benefits to humankind. They provide forest products and employment by means of forest industries like processing and trading of forest products. They provide landscapes for high recreational and spiritual value. Some of the other benefits include grazing, hunting, game management, protective functions (protection of soils, water resources, regulating micro and macro climate etc.) and nature (biodiversity) conservation and preservation. The carbon sequestering benefits of forests make them vital for the survival of human race in the long run.¹ Recently several studies have highlighted the potential use of wood for energy. In light of these benefits and increasing demand it is important to understand the difference between inventory and available inventory or potential supply.

As stated earlier the potential supply/availability of timber inventory to meet these demands is dependent on various factors. Specifically the biophysical characteristics describe the quantity, quality, and composition of the resource and the natural setting in which it exists. The economic and social factors determine whether the timber and potential goods

and services associated with it are desirable at their given economic, social and environmental costs.

Accordingly, the forest inventory availability can be categorized as:

1. **Biophysical availability**: The total amount of biomass/wood not constrained by biophysical constraints e.g. productivity, slope, physiographic class, water etc.

2. **Legal/Environmental availability**: The sum total of biomass available at different thresholds of biomass retention, as well as exclusion of areas with management objectives that in most cases exclude timber harvesting (e.g., wildlife management areas, scientific and natural areas, state and national parks, and wilderness areas). An important example of environmental constraint to available inventory is Riparian buffers.

3. **Economic availability**: The amount of inventory not constrained due to the economies of scale issues, higher opportunity costs (value of forest land in alternate uses), higher harvesting costs (steep slopes) and higher transportation costs (due to the distance from established roads).

4. **Social availability**: The amount of forest biomass that is not socially constrained and is otherwise compatible with landowner management objectives.

From the above definition it is evident that the question of available inventory can be effectively answered only by investigating the biophysical and economic characteristics of a region.

The general objective of this study was to examine various factors affecting the potential timber supply in North Carolina. Specifically, the study had two main objectives:
1. Econometric/Cross-sectional analysis of the biophysical and economic factors which affect the plot level forest harvest probabilities in North Carolina.

2. To investigate the impact of biophysical and economic factors on potential wood supply/availability in North Carolina.

The subsequent sections of the thesis comprise of three chapters. Chapter 2 reviews literature relevant to the present study and is further divided into three sections. In the first section literature related to theoretical modeling of timber harvesting has been discussed. Sections two, and three of chapter 2 discussed about research literature related to empirical models, and wood availability, respectively. Chapter 3 elaborated on the methodology, empirical model, and data sources used in this study. In chapter 4, empirical results from the econometric model and their consequences for the potential wood supply were illustrated. This chapter also included a discussion section for the results. Main conclusions from the study were listed in chapter 5.
CHAPTER 2 - REVIEW OF LITERATURE

2.1 Theoretical Models

The origin of most economic timber supply models is the Martin Faustmann model (1849). He showed that the value of a forest can be expressed as a sum of discounted net cash flow over an infinite time period. A forest owner's goal is to choose rotation so that the value of a forest is maximized.

Faustmann proposed the soil expectation value equation (Equation 2.1). Financially optimal rotation length is the point at which the marginal benefits from delaying harvests are equal to marginal costs of delaying harvest. Marginal costs (MC) of delaying harvest include forgone interest payments and value lost from delaying the future rotations. The marginal revenue include the value of the period by which harvesting was delayed, which would be equal to the growth of the period multiplied by the timber price.

Mathematically,

\[
Max \text{ } SEV = \frac{S(T) - w(1 + r)^T}{(1 + r)^T - 1} \quad (\text{Equation 2.1})^2
\]

Where \( SEV = \) soil expectation value,

\( S(T) = \) stumpage value at time \( T \),

\( w = \) planting costs,

\( r = \) discount rate,

\(^2\) http://foper.unu.edu/course/?page_id=167
\[ T = \text{harvesting age.} \]

\[
\frac{\partial S(T)}{\partial t} = S(T)' = rS(T) + rSEV \]

Where \( S(T)' = \text{Marginal revenue at harvest age.} \)

Some of the assumptions of the Faustmann model that made it unrealistic were steady and certain prices, costs, volumes, even-aged management, constant land use, etc. The other limitation of the Faustmann model was that it included only financial considerations.

Hartman (1976) proposed a model to incorporate amenity values. He incorporated a continuous stream of values, interpreted as recreation values, into the basic Faustmann model. He concluded that, as long as the amenity curve has a positive slope, inclusion of amenity values increases the optimal rotation lengths and can cause harvesting to be uneconomical.

Suppose \(^3\) that the private landowner values both the Faustmann revenue and the amenity services from forest. Based on Hartman (1976) the following quasilinear objective function can be postulated in the absence of taxes

\[ W = V + E \]

(Equation 2.3)

In equation (1) the first term is the Faustmann part, defined in the absence of taxes as

---

\(^3\) Optimal Forest Taxation under Private and Social Amenity Valuation.
Erkki Koskela* and Markku Ollikainen
The second term, \( E \) represented the net present value of amenities over all rotations.

The optimal rotation literature has emphasized landowners who are expected to maximize profits, for example forest industry landowners. Although the Hartman model seems more applicable for both nonindustrial and public landowners, it is a normative model with enormous data and assumption requirements. Hence, the benefits of applying the Hartman model are limited, while the costs are high. Hartman model was used as a base for various utilitarian studies such as Butler (2006).

\[ V = \frac{S(T) - w(1 + r)^T}{(1 + r)^T - 1} \]

\(^4\)Binkley (1981) used owner’s utility function to examine NIPF harvesting behavior. He differs from Hartman (1976) in a sense that according to him monetization of amenity values was not necessary. The subsequent empirical studies (e.g., Kuuluvainen et al. 1996; Dennis 1989; Butler 2006) after Binkley (1981) approved his utility function methodology.

**Link between Faustmann/Hartman models and Timber Supply:**

Faustmann and Hartman provide a theoretical basis for stand level harvest decisions. On the other hand the timber supply depends partially on plot decisions, but also includes the market and volume consequences of landowners’ aggregate response. For example, as described in the previous section, if the timber price increases the optimal rotation length decreases (from equation 2.2, the value of future rotations goes up, which is an opportunity cost). As the rotation length decreases the timber supply per acre will also decrease because

\(^4\)Favada (2009)
the timber will be harvested before reaching its maximum output years. On the other hand
with increase in prices more land will come under timber management, (because of the
decrease in opportunity cost) that will increase the timber supply. So the final impact on
regional timber supply would be uncertain. The above discussion demonstrated that
Faustmann’s optimal rotation is plot harvest description not a supply explanation.

Yin and Newman (1995) stated that the theoretical timber supply studies have tried to
convert the plot level Faustmann model to forest level supply model but produced false
results due to wrong assumptions. They explained that the timber age should not be
predetermined at the start of the rotation; rather it should be flexible and depend upon the
dynamic market conditions. They stressed that a profit function must include multiple
products and all kinds of realistic costs (regeneration, intermediate or operating, harvesting
costs etc.). In other words they tried to improve the timber supply formulation by
constructing a profit function starting from the timber production function.

Wear and Newman (1993) compared the profit functions of non-industrial private
forest owners (NIPF) and industrial private forest owners. Their results showed that there
behavior was consistent with a profit objective, but the NIPF production decisions are
governed by additional non timber values which affect the timber outputs from their land.
The authors concluded that NIPF put higher shadow values on their timber and other forest
assets than the industrial owners.

Wear and Park (1994) defined the available inventory in a region as amount of
timber available from all the ownership classes, for all management types, from all sub
regions considering the value of economic variables (harvesting, regeneration and transportation costs and product prices) in that region.

The available inventory (for an ownership of size L) in an empirically tractable way can also be defined as (from Wear and Park 1994).

\[
S_t = L \int_{q^-}^{q^+} \int_{\bar{a}}^{a^+} v(a, E; q) \phi(a, q) \, da \, dq = g(p; \phi(a, q))
\]

Here,

\( V(a, E; q) = \) per unit volume which is a function of age \( a \), management effort \( E \) and land quality \( q \).

- Here, \( S_t = \) Available inventory at time \( t \).
- Land quality varies from \( q^- \) to \( q^+ \)
- Age of standing timber varies from \( \bar{a} \) to \( a^+ \)
- \( \phi(a, q) = \) Density function which gives the relative frequency of land of quality \( q \) that is occupied by trees of age class \( a \).

### 2.2 Empirical Models

Binkley (1987) article illustrated the major paths of research which evolved over time in the field of economic modeling of timber supply. The author stated that the existing supply models do not do justice to the ecological information that is available. The paper emphasized on 4 major types of timber supply models: Long run models, short run models, transition models and household production models. The timeline for long run models can range from less than a decade to more than a century. Because of the lengthy time line these models allow the capital to adjust to economically desirable level. As per the author, long run
models are historically important and theoretically sound but their assumptions make it hard to address important questions of resource dynamics and short run impacts due to volatile economic conditions (for example short run development of timber prices).

In short run models, supply of timber are estimated by examining the impact of prices, inventory and other important variables (for example discount rate). The author stressed that although the models included in short run category are empirically tractable but these are not theoretically as sound as the long run models. His argument for theoretical weakness of short run models was justified. The limitations of the short run are due to the unique production function of timber and presence of multiple age categories. In addition to their theoretical weakness there are data issues with these models. For example in US there are no authentic data series (time series) for annual timber prices and inventory. Therefore he concluded that empirical models are constrained by lack of available data. The third types of models he discussed were transition models. The assumptions for these models were same as for the long run models. These models link supply in one period to forest growth and supply in another period. These models are powerful in a sense that they link short and long run supply behavior however the assumptions of perfect anticipation (of supply and demand by the owner) and demand structure limits their acceptance.

The fourth type of models Binkley called the household production models; consider non timber aspects of forests (non-forest income, education and age of the owner etc.) in determining the timber supply. The advantage of these models is that they include the multiple objective (e.g. profit, Leisure etc.) nature of the forest owners in studying their harvest behavior and timber supply. But the limitation of these models is unavailability of the
detailed information on landowner and ownership characteristics. These models primarily rely on forest owner survey data. In the United States, the USDA Forest Service's Forest Inventory and Analysis program is an extensive survey of forest biophysical attributes. Alternative data sources are required if economic variables are to be examined. There is a problem of aggregation of timber supply from an individual owner to an entire forest region in these types of models. The author emphasized the importance of improving data series for prices, harvest quantities and ownership characteristics by asserting that a unit spent on getting better data series would be more beneficial in terms of modeling supply than a unit spent on estimation techniques.

Wear and Parks (1994) presented a theoretical framework (for aggregate timber supply) and used that to illustrate the existing literature for timber supply models. They categorized the models in existing literature as those that focus on individual harvest rules (Normative models and Positive models) and those that focus on aggregate timber supply (engineering models and econometric models). Normative models are based upon the bare land approach (Faustmann’s Model). These models can be formulated as a dynamic optimization problem (with sequence of rotations as decision variables). As per the article, the major drawback of these models was that they ignore the important aspects of the problem by assuming a regulated (forest age between zero and optimum rotation age) forest, and constant timber production. The authors explained that even with their limitations normative models are good for developing insights into the timber supply.

In the article, empirical forms of decision models were illustrated as “positive harvest models”. In positive harvest models utility depends upon vector of stand attributes (age,
management intensity, land quality) and owner characteristics such as owner income, education etc. The optimum harvest rule depends upon marginal utility of delaying harvest and marginal cost from delaying harvest. The examples of these models are harvest choice models (probit, logit and tobit models). The engineering models were constructed from normative individual harvest models. The forest land base was arranged into different categories (based upon the quality attributes) and harvest quantity for each quality class was determined. The authors stated that these models are authentic because of their consistency with the capital theory but their effectiveness is limited by their long run focus.

Normally the long run is so far in forestry as to be inappropriate to the current problems at a time. These models are also inclined to specification error (because these are not based upon historical observations). The next categories of aggregate supply models illustrated in the article were econometric timber supply models, which are effective tools for testing economic hypothesis. In these models, the supply is modeled as timber price, capital inputs (timber inventory, effort, land quality etc.) and other supply shifters. The author called these models empirically thorough but theoretically weak (due to lack of explicit links to the production technology). The authors suggested that, an improvement in timber supply studies can be done by linking long term models with economic decision models.

Dennis (1989, 1990) used a household production model in which he integrated forest, owner, and economic variables for an economic analysis of harvest behavior. The tobit analysis results in (Dennis 1989) his article illustrated strong relation of harvest behavior with forest characteristics (e.g. timber inventory, species) and land owner characteristics (e.g. Income, education, occupation etc.). But prices, which are an important
economic variable, were found to be non-significant for harvest behavior (the author assumed the non-responsive ness of prices was due to opposing substitution and income effects and multicollinearity or measurement error).

Wear et. al (1999) investigated the impact of population growth on timberland area and timber inventories. They used US census data, expert opinion maps and FIA data to study the effect of population density on timber supply. The logistic regression model was used to investigate the impact of population density on commercial forestry. The interesting result they found was, the commercial timber management has a negative relationship with the population density and as the population density reach 150 per square mile the probability of commercial forestry become zero.

Prestemon and Wear (2000) used a “bottom-up” type of approach that provided a useful means of defining the market implications of changes in supply factors. They argued that aggregate supply models which have typically estimated as single equations for broad regions and broad owner groups (e.g., forest industry, nonindustrial private forests) provide a useful structure for examining the short-term implications of changes in timber demand but these models, as usually specified, do not provide especially useful insights into the consequences of changes in the supply structure of timber. This shortcoming is the result of the aggregation process implied by aggregate supply models. Aggregate supply at the regional level is defined as a response to changes in aggregate inventory quantity. Any change in the structure of forests (e.g., urban development) must therefore be channeled through its effect on aggregate inventory to define an impact on total timber produced. An alternative approach is to explicitly aggregate the outcomes of harvest choices made at the
micro (forest stand) level. In their study they investigated aggregate timber supply by ownership for a small region by applying stand-level harvest choice models to a representative sample of stands and then aggregating to regional totals using the area-frame of the forest survey. Timber harvest behavior were estimated using probit models for three ownership categories (NIPF, industry, and government) in coastal plain southern pine stands of North Carolina using individual permanent and re measured stand-level data from FIA surveys. The timber harvest decision was modeled as a function of timber values, a cost factor, and stand volume as a proxy for non-timber values. Probit models were statistically significant at 1% for all ownerships. Area expansion factors (the portion of forest area in the region represented by the sampled stand) were then combined with harvest probabilities to model the aggregate effects of price changes on timber supply, given a fixed forest area.

Pattanayak et al. (2002) in their study tried to fill some gaps in the existing literature by explaining the relationship between the characteristics of forest owner and timber supply in terms of the optimal rotation problem (thereby combining the two fundamental streams of modeling, one that stresses on optimal harvest age and other which focuses on the effect of owner characteristics and timber supply constraints). They used three stage least square (3SLS) procedure to generate theoretically sound parameter estimates for timber supply functions.

Polyakov, Wear, and Huggett (2010) explicitly linked aggregate timber supply models to observations of individual harvest behavior. They used harvest choice models to derive complete aggregate supply models for a broad region. They used various site characteristics (type of soil, slope, and species), location (physiographic region) and
economic variables to model the probability of harvest for individual plots using forestry inventory (FIA) data base.

Bolkesjø et al. (2008) have done econometric analysis of saw log and pulp wood supply with respect to saw log and pulp wood prices. They tested time series, cross-sectional and panel data models. The data used was from 102 Norwegian municipalities, observed from 1980 to 2000. Their results exhibited positive own price elasticities among saw logs supply and saw log prices and negative elasticity with pulpwood prices, but the pulpwood supply had positive elasticity with both saw log prices and pulpwood prices.

2.3 Wood Availability

Numerous timber supply studies have generally used econometric approaches to model either aggregate or individual harvesting behavior of specific types of forest owners, such as industrial owners (Newman and Wear 1993) or family forest owners (Binkley 1981). In all these studies, land characteristics, landowner characteristics, and various economic and factors are correlated with the timber supply. A very few studies have used the econometric estimates to determine the availability of timber in a region. One such study is Butler et al. (2010). In this study the social and biophysical availability of the wood located in family forests across the northern United States was investigated. The researchers concluded that the biophysical, social and ownership related constraints reduce the available inventory by nearly two-thirds. Majority of reduction was due to the owner attitudes. But the major drawback of the study was the use of availability thresholds on the basis of expert opinions and existing literature estimates rather than empirical model. They used parameter estimates (to formulate
thresholds for availability) from previous research and harvesting guidelines (from state manuals) to formulate their wood availability model. For example they used 50\% threshold for slope based upon Kittredge and Parker (1999) study (Massachusetts and New York states stipulate harvesting restrictions on slopes above 60\% and 30\% respectively). They also used a 0.75 threshold for availability, which means that if the plot is constrained for wood availability because of one constraint (slope), it reduces the wood availability by 0.75 times. The 0.75 reduction was based upon expert opinion not any scientific model.

In the current study, to make the potential supply / availability empirically justifiable, probability graphs based upon the econometric analysis (logistic regression) were presented and thresholds to separate available inventory from unavailable inventory were discussed.

2.4 Summary

Theoretical timber supply modeling came a long way from Martin Faustmann (1849) to Hartman (1976) to Binkley (1981) and the other modern attempts in last thirty years that includes Newman and Wear (1993), Wear and Parks (1994), Yin and Newman (1995) etc. Researchers applied innovative techniques (restrictive profit function for short run supply by Newman and Wear 1993) to improve the understanding of the harvest behavior and regional timber supply. There have been attempts to link Faustmann’s optimal rotation stand based model to regional timber supply from heterogeneous ownership categories.

There was also very interesting development in the empirical literature, for example the individual harvest decision models developed using the ownership and stand characteristics. Some individual harvest based studies tried to link individual estimates to the
aggregate regional supply. For example “bottom up” approach used by Prestemon and Wear (2000) to estimate supply in coastal region of North Carolina. There are some studies (Pattanayak et al. 2002 used three stage least square estimator to generate theoretically sound parameters) which used sophisticated econometric tools to try to link the theoretically sound optimal harvest approach to empirically more manageable models.

Analytically, harvesting was treated as a discrete choice in most of the individual-choice models. The general analytical approaches used to test econometric relationships between timber harvesting and various site characteristics, ownership characteristics and financial variables are logistic (Jamnick and Beckett 1988, Hyberg and Hoithausen 1989), probit (Dennis 1990, Prestemon and Wear 2000, Pattanayak et al. 2003) and tobit (Dennis 1989) models. These econometric models provide very powerful tools for checking various economic hypothesis but these are not explicitly linked to the theories of production behavior (Binkley 1987, Wear and park 1994).
CHAPTER 3 – METHODOLOGY AND DATA

3.1 Introduction

Following Wear (1994), the current study uses a positive harvest model based upon individual harvest decision making to econometrically test the impact of various biophysical and economic variables on harvest and wood availability (potential supply). The theoretical model presented here was a modified version as postulated by Wear and Park (1994) and used by (Polyakov et al. 2010).

3.2 Theoretical Model

The standard empirical characterization of timber supply is of the form (Wear and Park 1994):

\[ Q_{it} = f(P_{it}, I_{it}, Z_{it}) \]

(Equation 3.1)

Where:

- \( Q_{it} \) = quantity harvested in region \( i \) at time \( t \)
- \( P_{it} \) = price of stumpage or logs in region \( i \) at time \( t \)
- \( I_{it} \) = standing inventory in region \( i \) at time \( t \)
- \( Z_{it} \) = a variety of other factors including interest rates and landowner characteristics.

The given model was derived from forest production function implicitly. There should be positive sign for price because higher the stumpage price more is the economic
availability of timber inventory. Also, higher stumpage prices make timber harvest relatively more valuable than non-timber benefits of forests (e.g. amenities). Timber inventory has positive impact on harvest because at higher inventory harvesting costs are lower. Also more the inventory lower is the impact of harvest on non-timber values (Binkley 1987). There can be a number of supplier shifters which are candidates for inclusion in Z (discount rates with positive sign, ownership characteristics for example income, profession, education etc.).

The above timber supply formulation is widely used in forest sector market analysis. It is not perfect but is certainly defensible as it is derived from forest production function. The production function of timber as a commodity is different from other (industrial or consumer) goods in a sense that it takes a long time to grow it to merchantable size and quality. Therefore its production function includes planting costs and years of maintenance costs, for example land rent, labor, management, fertilization, thinning etc. and once the product is ready generally it is not harvested all at once. In the given characterization, In equation 3.1, production function is implicitly represented by Inventory (*I*_it). This is because of the reason that the inventory represents the productivity of the soil, years of costs and hard work spent on the stand. Therefore the inventory can be characterized as:

\[ I_{it} = f(a,k,e) \]  
\[ (Equation 3.2) \]

*Where k = Land quality

* e = various types of costs (management effort, land rent, labor, seedlings, planting cost fertilizers, thinning etc.)
\[ a = \text{stand age} \]

Once the timber is of merchantable size and quality, generally it is not all cut at once. The amount of harvest depends upon various factors for example timber prices, which is represented by \( P_{it} \) in equation 3.1. At the time of timber supply/harvest various opportunity costs (interest rates, cost of land in other uses etc.) are represented by \( Z_{it} \).

From the discussion it can be concluded that above definition of timber supply is defendable. But there are some limitations which are due to the unique production function of timber and poor quality of data availability, such as ownership data, product prices data etc. Age structure of the stand is very important variable to determine timber supply. The misspecification error may be serious when the age structure and species composition of the forest is not constant (Wear and Park 1994). Also the ownership objectives are very important, Pattanayak et al. (2002) illustrated that the objectives for the various ownership categories may differ in a sense that timber harvesting decisions also depends upon non timber amenities (for example recreation, hunting) are correlated with the structure of forest capital, ownership, and management.

**Proposed Model**

With the given limitations of the data availability and unique life cycle of the timber, the following supply function was proposed:

\[ Q_i = f(P_i, I_i, Z_i, L_i) \]  
(Equation 3.3)
The subscript t was deleted as there was only one time dimension. $P_i$ represents the timber product prices in the market near a timber plot $i$. $I_i (a, k, e)$ represents the inventory or volume at a given plot $i$. It was formulated as a function of stand age, land quality and management effort. In addition to the three original variables, one more variable of $L_i$, large tree proportion in the plot $i$ was added to capture age distribution of the stand. Age distribution is of immense importance for timber supply. $Z_i$ represent the opportunity costs for timber land (interest rate, price of land for development etc.). In the empirical analysis, variables of housing prices and economic tier of the county were used as a proxy for the opportunity cost. Theoretically this model is somewhat weak because it lacks important variables (as per economic theory) like vector of all major costs ((planting, harvesting, transportation, processing (traditional and bioenergy related also)), and ownership characters.

The production function for timber supply was presented in equation 3.4. The inputs included, age of the forest, $a$, level of forest management work, $W$, land quality, $k$ and large tree proportion, $L$.

The merchantable timber volume per unit area, $V_i$ is given by the production function,

$$V_i = q(a, W, L; k)$$

(Equation 3.4)

The marginal physical products of $a$, $k$, $W$ and $L$ are all increasing at a decreasing rate.
Given that the forest manager’s objective function and discount rate can be specified, the forest production function can be used to define whether and when a forest stand would be harvested. For example, consider a manager, who faces prices $p$ for timber and $W$ for management work effort.

When the land is maintained for forever in forest use for profit, the manager will maximize profit by selecting harvest ages and levels of effort, $W$, to optimize:

$$\pi_0 = \max \sum_{j=0}^{\infty} \{pq(a, W, L; k)e^{-ra} - wZ\}e^{-ra}$$  \hspace{1cm} \text{(Equation 3.5)}$$

$\pi_0$ = optimum profit obtained (which is the present net value for an infinite sequence of identical harvest ages)

$p$ = timber price

$W$ = management work effort

$L$ = proportion of large trees

$w$ = unit opportunity cost

$Z$ = opportunity cost of land

$k$ = land quality

$r$ = the interest rate
Based upon the Faustmann notion the land owner or the firm will harvest when the marginal cost (opportunity cost of land) of delaying the harvest will be equal to marginal benefit of delaying the harvest.

If the current price level is \( p \), manager’s optimum harvests age, \( a^* \), is given by:

\[
a^*(p;q) = a : MB (a, W, L; k) = MC (a, W, L; k), \quad \text{Given } MC>0, \quad \text{(Equation 3.6)}
\]

Where, \( MB = \) marginal benefits of delaying harvest,

\( MC = \) marginal opportunity costs of delay,

Optimum harvest age \( a^* \) depends on market prices \( (p) \). This optimum age is not necessarily the same as that given by the timber only (no other products like amenities considered) solution and may vary over time as prices, \( W \) and \( Z \) change. Also, the relationship between \( MB \) and \( MC \) need not be viewed as strictly deterministic or static and would be dynamic as risk and price preferences of the land owner. For example \( MC \) may change with time and region. \( MC \) of timber production in urban areas is more as the land prices for development are more than rural areas. The conclusion of the above discussion is that, once the expected marginal returns to delaying harvest are no longer greater than the marginal opportunity costs of delaying the harvests, the harvest age, volume and effort are defined. This then can be used as a two-period model where as long as \( MB > MC \) for delay between the two periods, then harvest is deferred. Otherwise, harvest occurs.
A two-period model is implied where harvest occurs ($H = 1$) when the MB is equal to or less than the MC for a forest plot where these values depend exclusively on the attributes of the plot and the ability to forecast end-of-period values,

$$H = \begin{cases} 
1, & \text{if } MBD(q) \leq MOC(q) \\
0, & \text{otherwise}
\end{cases} \quad \text{(Equation 3.7)}$$

(Polyakov et all. 2010)

The decision variable in this formulation is, whether or not to harvest at the beginning of the analysis period) and depends on the benefits and opportunity costs of harvesting. It therefore depends on the ability to estimate net harvest benefits for the two periods being analyzed.

### 3.3 Analytical Model

**Binary logistic regression**

In this study binary logistic regression model was used to estimate the individual harvest behavior (harvest probabilities). Binary logistic regression is a form of logistic regression which is used when the dependent is a dichotomy and the independents are of any type. This is a common approach used in the past for modeling timber harvesting behavior (Prestemon and Wear 2000, Butler 2006). Logistic regression can be used to predict a dependent variable on the basis of continuous and/or categorical independents and to
determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents; and to understand the impact of covariate control variables. The impact of predictor variables is usually explained in terms of odds ratios.

In this type of regression, maximum likelihood estimation is used after transforming the dependent variable into a logit variable (the natural log of the odds of the dependent occurring or not). In this way odds of a certain event occurring are estimated. The procedure calculates changes in the log odds of the dependent, not changes in the dependent itself as OLS regression does.

To transform the discrete outcomes to a continuous scale and obtain many of the desirable features of ordinary linear regression, the log of the odds is calculated (Equation 3.8). Maximum likelihood techniques are used to solve the following equation.

\[
\text{logit } (p) = \log \left( \frac{p}{1-p} \right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k = \alpha + x \beta \quad \text{------------------} (\text{Equation 3.8})
\]

The above logistic regression model describes a linear relationship between the logit, which is the log of odds, and a set of predictors.

Where, \( p = \) probability of the event,

\( \left( \frac{p}{1-p} \right) = \) Odds of the event occurring,

\( \alpha = \) intercept term

\( \beta = \) vector of Coefficients
X = vector of independent variables.

The model can be interpreted using the logit scale, or the log of odds can be converted back to the probability such that:

\[
p = \frac{\exp(\alpha + x \beta)}{1 + \exp(\alpha + x \beta)}
\]

(Equation 3.9)

The advantage of using the logit scale for interpretation is that the relationship between the logit and the predictors is a linear relationship.

In the present study dependent variable has two categories. The final model was selected using an iterative process (as in Butler 2006). There were two steps used to select the variables for the final model:

**Bivariate\(^5\) logistic regression model:** In bivariate logistic model all the independent variables (biophysical, economic and location) were analyzed separately to investigate their relationship and predicting power with the dependent variable (harvest probability).

**Multivariate\(^6\) logistic regression model (complete set of predictors):** After testing for their individual relationship with the dependent variables a multivariate model was used to examine any statistical issues (multicollinearity etc.)

---

\(^5\) A procedure for deriving the equation that relates a single dependent variable and a single independent variable.
**Multivariate logistic regression model (final):** Final multivariate model was developed by accounting for the multicollinearity issues in the model in step two. The predictor variables in the final model were selected based upon their explanatory power (pseudo $R^2$) and theoretical importance (theoretical model).

**Goodness of fit of the model**

There is no goodness of fit statistic in logistic regression such as $R^2$ in ordinary least squares (OLS) regression. The maximum likelihood estimates from a logistic regression arrive at through an iterative process. They are not calculated to minimize variance, so the OLS approach to goodness-of-fit does not apply. However, to evaluate the goodness-of-fit of logistic models, several pseudo $R^2$ exist in literature. These are called "pseudo" $R^2$ because they look like $R^2$ in the sense that they are on a similar scale, ranging from 0 to 1 with higher values indicating better model fit. The goodness-of-fit and predictive power of the models in the present study was assessed using log likelihood ratios and McFadden's Pseudo $R^2$. The log-likelihood ratio quantifies the strength of the model compared to a model with no predictor variables; the associated $X^2$ statistic was used to assess the significance of this statistic. The formula for McFadden's Pseudo $R^2$ is as follows:

$$Pseudo\ R^2 = 1 - \frac{lnL(M_{full})}{lnL(M_{intercept})}$$  \hspace{1cm} (Equation 3.10)

---

6 A multivariate statistical model is a model in which the effect of more than one independent variables on the dependent variable is modeled simultaneously.

7 [http://www.ats.ucla.edu/stat/mult_pkg/faq/general/psuedo_rsquareds.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/psuedo_rsquareds.htm)
\( M_{full} \) = model with predictors,

\( M_{intercept} \) = model without predictors,

\( L \) = estimated likelihood.

The Pseudo \( R^2 \) value tends to be smaller than \( R \)-square and values of 0.2 to 0.4 are considered satisfactory.

### 3.4 Logistic Regression Variables

**Dependent Variable**

In the current study dependent variable (harvest) was formulated as a dummy variable having two categories as characterized below:

\[
y = \begin{cases} 
0 & \text{if no harvest} \\
1 & \text{if harvest} 
\end{cases} \quad \text{(Equation 3.11)}
\]

FIA plot based measurements were used to formulate the dependent variable. Plots where the removals were greater than twenty five percentage of the pre harvesting volume or harvest intensity was greater than twenty five percentage (Equation 12) were categorized as harvested and assigned a harvest value of one. All other plots were allotted a harvest value of zero. Approximately twelve percent of the forest plots in North Carolina were classified as harvested over the 2003-2010 period.
\[ H = 1 \text{ if } \left( \frac{\text{Removal Vol}}{\text{Preharvest Vol}} \right) \times 100 > 25 \% \text{ of Preharvest Volume} \quad \text{(Equation 3.12)} \]

Twenty five percent threshold was specified on the basis of existing literature (Butler 2006 & Prestemon and Wear 2000). The threshold appeared reasonable, since using this threshold, 94 percent of the removal volume was from the harvested plots.\(^8\) The box plot (figure 2.) for the harvest intensities proved the adequacy of the threshold. The box plot demonstrated that, using above threshold criteria, 75 \% of the plots from where removal volume was greater than zero, were designated as harvested. These 75\% of plots represented 94\% of the removals, so the assumption seemed reasonable.

**Independent Variables**

The Independent/ predictors variables were categorized into economic/financial, biophysical & Location categories.

**Economic/ Financial Variables**

Forest product prices, stumpage values, economic tier of the county and housing prices were used as economic predictor variables. Descriptions and summaries of economic variables used in the empirical analysis were presented in table 3.

Timber product price data (pine saw timber, pine pulpwood, chip-n-saw, hardwood pulp and hardwood saw timber) was obtained from Forest2market Inc. The product prices were expected to be positively correlated with the probability of harvest.

---

\(^8\) Butler (2006), used a box plot to specify threshold for the dependent variable.
To investigate the impact of opportunity cost of timber production on harvest probabilities a dummy/categorical variable for the economic status of the county was used. The assumption behind this variable was that in a more developed county (Tier 3) the developmental prices of the land are more and hence the harvest probability should be low due to the conversion of timber land to the developmental land. The county wide housing prices were also used as a continuous variable to test the impact of opportunity cost.

**Biophysical Variables**

The present condition and expected future condition of a forest stand depends upon its biological and physical attributes. To investigate the impact of biological and physical variables on harvest probability the predictor variables were derived from FIA forest inventory data. The biophysical predictors included in this study were softwood volume, hardwood volume, softwood proportion, number of large trees, proportion of large trees, land productivity, physiographic class (xeric, mesic and hydric), management classes and slope. The descriptions and summaries of biophysical variables used in the empirical analysis of timber harvesting were presented in table 1 and 2.

Volume variables were calculated for softwoods (SV) and hardwoods (HV) to assess the relationship of these species groups to harvesting behavior. Softwood proportion (SPROP) variable was also formulated from the plot level estimates of FIA data. The use of SPROP as one of the predictors was based upon the assumption that traditionally softwoods command higher market values and have larger markets. Therefore softwoods should be more strongly and positively correlated with the harvest probabilities than the hardwoods.
Hardwoods may increase amenity values due to aesthetic appeal and increased biodiversity. Two other biophysical variables used in the analysis were number of \(^9\)large trees and proportion of large trees on the plot. These two variables were used to represent the trees which are of harvestable size or age. The hypothesis was that if the proportion and number of large trees was more the probability of harvest would be more. These were used as a proxy for age distribution as the actual age distribution could not be estimated due to the complex nature of FIA data.

The forest management classes used to examine their impact on the harvest probabilities were: planted pine, natural pine, oak-pine, upland hardwood, and lowland hardwoods. Dummy/categorical variables for all management classes were used with planted pine as base case.

The impact of land productivity was examined by including a categorical variable for productivity. FIA \(^10\)classified (SITECLCD variable) forest land in terms of inherent capacity

\(^9\) As per the FIA definition large tree, or sawtimber stands include softwood tree at least 9.0 in. dbh (diameter at breast height) or hardwood tree at least 11.0 in. dbh. Medium, or poletimber, stands as forestland dominated by softwoods between 5.0 and 9.0 in. dbh or hardwoods between 5.0 and 11.0 in. dbh and small (seedling/sapling stands) forestland dominated by trees less than 5.0 in. dbh.

\(^10\) SITECLCD variable in condition table of FIA data.

Code Description:

1. 225+ cubic feet/acre/year
2. 165-224 cubic feet/acre/year
3. 120-164 cubic feet/acre/year
4. 85-119 cubic feet/acre/year
5. 50-84 cubic feet/acre/year
6 20-49 cubic feet/acre/year
7. 0-19 cubic feet/acre/year
to grow crops of industrial wood. They identify the potential growth in cubic feet/acre/year which is based on the result of mean annual increment of fully stocked natural stands. More productive land was assumed to be positively correlated with the harvest probability.

Dummy/Categorical variables for productive (capacity to grow more than or equal to 50 cubic feet/acre/year) and non-productive (capacity to less than 50 cubic feet/acre/year) land were formulated.

The general effect of accessibility and moisture availability to trees was investigated by using physiographic class variable. In FIA data the physiographic class variable has 3 categories. The sites with normally low or deficient in available moisture are called xeric (dry tops, dry slopes, deep sands etc.), sites with normally moderate but adequate available moisture are called mesic (flat woods, rolling uplands, moist slopes, flood plains) and sites with normally abundant or overabundant moisture all year are called hydric (swamps, small drains, bays, wet pocosins, ponds etc.). The harvest probability in hydric sites was assumed to be lower than mesic and xeric classes because of the problem of inaccessibility and higher harvesting costs.

Slope (in percentage) was used as a continuous variable as given in the FIA data. Steep slopes were hypothesized as a negative factor for the harvest probabilities as forests located on steep slopes are less operable and therefore less likely to be harvested.

**Location Variables**

Physiographic regions (figure 3) in North Carolina are northern coastal plain, southern coastal plain, piedmont and mountains. These four regions present diverse
conditions in terms of weather, productivity, harvesting, market, urbanization and management classes. To investigate the interregional impact of these factors on harvest probabilities, the physiographic regions were used as dummy variables (southern coastal plain as base).

3.5 Data

With the above mentioned general theoretical and empirical framework, the impact of biophysical and economic variables on timber harvest probability and timber availability in North Carolina was investigated.

The amounts of forest inventory estimates (pre harvest volume, softwood volume, hardwood volume, removals) were calculated by using the data collected by the US Forest Service Forest Inventory and Analysis (FIA) program (Bechtold and Patterson 2005). FIA has established a set of stratified, random inventory plots across the United States to estimate, among other attributes, species composition, forest health, timber volume, and biomass. The United States was divided into hexagons of approximately 6,000 ac, and a randomly located sampling point was established within each hexagon. This point serves as the center for a permanent inventory plot. If any part of the plot is determined to be forested, based on remotely sensed imagery and field verification, a field crew will collect data on the trees and other site variables (US Forest Service 2005). Because each plot is randomly selected, the measured attributes can be used to estimate population-level statistics.

Each of US states has been surveyed multiple times. However, the dates of forest surveys are different across states. FIA use an annual approach (approximately 20% of plots...
are re-measured annually and all sample plots are re-measured on a 5-year cycle\(^{11}\)). In the current study for North Carolina the measurement cycle included years from 2003 to 2010.

Forest Inventory and Analysis data sets are stored in tables, three of which were utilized for the present analysis. The plot, condition, and tree tables (used in this study) provide information on the overall plot characteristics, discrete landscape features, and measures associated with individual trees larger than an inch in diameter, respectively. Each plot represents a larger portion of the landscape to estimate the total inventory—the representative area is called the expansion factor.

**Price data**

The price data was obtained from Forest2Market, Inc. Forest2Market is an online timber pricing service. Their price data is available at \(^{12}\)micro market level. Micro market is a smaller unit within a state. There are seven micro markets which cover North Carolina. The harvest choice models in the past (Polyakov et al 2010, Butler 2006 and Prestemon 2000) used TimberMart-South as their source for product prices. TimberMart-South report product prices at a region bigger than a micro market. There are two TimberMart \(^{13}\)regions per state. Forest2Market price data is more suitable for a cross-sectional analysis like the current study because of more regional variability. All the plots with in a micro market were attributed the same price. Because the FIA plot data set was spanned from years 2003 to 2010, therefore to match the plot harvest levels, average of product prices across all these years was used.

---

\(^{11}\) The measurement cycle can extend beyond 5 years due to budgetary constraints. For example here for North Carolina the measurement cycle length is 7 years.


\(^{13}\) [http://www.timbermart-south.com/report.htm](http://www.timbermart-south.com/report.htm)
Other Economic Data

The county economic tier designation data was obtained from N.C. Department of Commerce. The N.C. Department of Commerce annually ranks the state’s 100 counties based on economic well-being and assigns each a Tier designation. The 40 most distressed counties are designated as Tier 1, the next 40 as tier 2 and the 20 least distressed as tier 3. Economically most troubled counties of tier 3 were compared for harvest probability with tier 2 and tier 3 counties. The county vise housing prices were used as a continuous variable to test the opportunity cost for forest land. The average house listing price for North Carolina counties was obtained from Trulia, Inc.

---

14 The N.C. Commerce Department ranks counties by a formula that considers each county’s unemployment rate, median household income, population growth, and assessed property value per capita. Additionally, all counties with a population of less than 12,000 and any county with a population less than 50,000 where 19 percent or more of those people live below the federal poverty line, are automatically designated distressed or Tier 1 counties.
16 http://www.trulia.com/home_prices/North_Carolina/
CHAPTER 4 - RESULTS AND DISCUSSION

4.1. Introduction

The variables used for the econometric investigation of the harvesting behavior were selected on the basis of the theoretical model discussed in chapter 3. The model was estimated using 2003-2010 NC FIA data supplemented by price and demographic data as described in chapter 3. Since the FIA data is focused on biological inventory, implementation of the theoretical model is constrained. For example FIA data do not provide any ownership information such as type (Family, corporate etc.), age, education level, Income, size of land holding, product prices etc. These are very important variables in terms of studying the harvest behavior and wood availability. The current study was an attempt to estimate predicted harvest probability and wood availability given these constraints.

4.2. Bivariate Analysis

The bivariate models were used to investigate the individual impact of various predictor variables on harvest probability. Based upon the results of bivariate models multivariate model was developed (Butler 2006) and used to investigate the complex relationship of the predictors with harvest. The final multivariate model included only a subset of the variables because of the issues of multicollinearity and quasi-complete separation.

17 The meaning of quasi-complete separation is that, when the dependent variable variable separates an independent variable completely. Albert and Anderson (1984) define this as, "there is a vector α that correctly allocates all observations to their group."

http://www.ats.ucla.edu/stat/mult_pkg/faq/general/complete_separation_logit_models.htm
The coefficient estimate (table 4.) for softwood volume (SV) was positive and highly significant (at < 1%) which meant that with increase in volume, probability of harvest increases.

The odds ratio of 1.066 (table 4.) for SV variable meant that with every unit (100cuft/acre) increase in volume the odds of harvest over no harvest increases by 1.066 times and the relative probability of harvest increases by (1.066 - 1)*100 = 6.6%. The result was as expected from theoretical point of view because at higher volume per unit harvesting costs are lower. Also more the inventory volume lower is the impact of harvest on non-timber values (Binkley 1987).

The estimate for softwood proportion variable was positive and highly significant (p=.0001). The odds ratio of 3.614 suggested that with each unit increase in softwood proportion the probability of harvest increases by 261.4% or 3.614 times. This result proved the economic importance of softwood. The results also demonstrated that with a unit increase in large tree number the probability of harvest would increase by 1.009 times and with every unit increase in large tree proportion (percentage) the probability of harvest would increase by 3.56 times.

In the results for economic variables the estimate for pine saw timber price was positive and highly significant. The result was expected because pine saw timber is the most valuable product attained from the final harvest of timber. Also, higher price makes timber harvest relatively more valuable than non-timber benefits of forests

---

18 A log or tree that is large enough (usually 10 to 12 inches in diameter) to be sawed into lumber. Minimum log length is typically 8 feet. http://www.ces.ncsu.edu/forestry/pdf/WON/won26.pdf
(e.g. amenities). The odds ratio suggested that with every unit increase in the softwood price the probability of harvest increases by 13.9%.

Pine pulpwood (PP) is also an important commercial product from timber harvest. The estimate for PP prices was positive and statistically significant (p=.0001). The odds ratio of 1.356 confirmed that with every unit increase in the softwood price the probability of harvest increases by 35.6%.

Among other products chip-n-saw is very important. Chip-n-saw mills provide a market for trees larger than pulpwood and smaller than saw timber. Theoretically, chip-n-saw prices should be positively related with the harvest probability. The empirical results from the logistic regression confirmed that with every unit increase in chip-n-saw prices the harvest probability increases by 16.1% (odds ratio=1.161).

The bivariate model suggested that the economic tier of the county, house prices, and average stumpage value in the county (in which the plot is located) affect the harvest probability significantly. The estimate for the economic tier suggested that if a plot is located in the economically higher tier county the probability of harvest decreases by 25.8% (odds ratio=0.742). The reason for the lower harvest probability in the more economically developed counties was the higher developmental prices (which is an opportunity cost for forest land).

Average housing prices were negatively correlated to the harvest probability. As per the results, with every unit (1000 dollars) increase in housing price, the harvest probability decreased by 0.3% (odds ratio=.997). This may be due to the reason that higher housing

---

prices increase the demand for land for constructing houses and hence displacing the timber land. House listing price variable was used as a proxy for opportunity cost. Studies in the past (Wear et al. 1999) suggested that population in a county impact the harvest probability negatively. The results of the binary logistic model in this study did not confirm that, while the coefficient was negative but statistically insignificant (p=0.5).

The estimates for the dummy variables for physiographic regions suggested that harvesting probabilities were low in mountain (mt) and piedmont (pd) regions as compared to southern coastal plain (SCP). If a plot is located in the mountain region its probability of harvest was 68.6% (odds ratio=0.314) less than a plot in SCP region. Similarly the probability of harvest for a plot in piedmont region was lower by 26.7% (odds ratio=0.733) as compared to SCP region. On the other hand if a plot is located in a Northern coastal plain region its probability of harvest was 32 % (odds ratio=1.320) more as compared to a plot in SCP. The result reflected that harvest is not equally distributed throughout the state of North Carolina. This is due to the reason that most of the planted pine plots (harvested and non-harvested) are located in coastal plains (figure 5). There are few in piedmont and very few in mountains. Planted pine is important because of its market importance. In piedmont region urbanization also impact the timber land negatively, which lower the harvest levels. The availability of wood in the long run will be strongly affected by the overall amount of forestland that is being lost to development across many parts of the United States. In fact, Stein et al. (2005) estimated that the United States loses 2,500 ac of forest per day to this largely irreversible change.
The results suggested that other plot variables such as productivity, physiographic class (hydric) and slope impact the harvest probability in a significant way. For the forest plots having good soil productivity (SITECLCD = 1 to 5) the probability of harvest increases by 164.6% (odds ratio=2.646). This is because of the reason that the plots which have low productivity are unlikely to be intensively managed.

The probability of harvest in a hydric region was lower by 59.9 % (0.401) as compared to a plot in a mesic region. The lower harvest probabilities in hydric sites were due to higher logging costs.

The estimate for the slope variable was expected to be negative, as the increase in slope results in more costly harvesting operations. The empirical results confirmed that, a unit increase in slope percentage results in a decrease of 2.6% (odds ratio=0.974) harvest probability.

The negative signs of the estimates for categorical variables for management classes (forest types) suggested that the harvest probability for planted pine plot (bmt=1) was more than natural pine (bmt=2) and oak pine ( bmt=3). The coefficients were not significant for upland hardwoods (bmt=4) and lowland hardwoods (bmt=5).

The pseudo $R^2$ for individual variables in the bivariate logistic regression results (table 4) demonstrated that among the bio physical variables softwood volume (with pseudo $R^2$=0.079), softwood proportion (with pseudo $R^2$ =0.041), number of large trees (with pseudo $R^2$=0.044), proportion of large trees ( with pseudo $R^2$=0.025),the dummy variable for mountain region, mt (with pseudo $R^2$=0.025), slope (with pseudo $R^2$ =0.018) and land productivity (with pseudo $R^2$ =0.01) were the most important in terms of explaining the
harvest probability. While, among the economic/financial variables product prices for pine sawtimber (with pseudo $R^2 = 0.017$), hardwood sawtimber (with pseudo $R^2 = 0.017$), chipnsaw (with pseudo $R^2 = 0.014$) and pine pulpwood (with pseudo $R^2 = 0.011$) were the most important (in descending order as per their $R^2$ values) to explain harvest probability. The lower pseudo $R^2$ results for all these predictors proved that separately all these variables explain very little of the variation in harvest.

### 4.3. Multivariate Logistic regression

The bivariate model was used to identify the important variables and investigate their individual impact on harvest probability. The results suggested that although these variables were important but their separate predictive power for harvest was low (low pseudo $R^2$ values). In the real world scenario all these variables affect the harvest probability simultaneously. Therefore it’s logical to study their impact in a multivariate way. An iterative process was used to build the multivariate logistic regression model. The iterative process was used to better isolate problematic combinations of variables. Initially all the variables of bivariate model were retained (table 5). Based upon the results from the multicollinearity and quasi-complete separation tests and the predictive power of the variables only a subset was retained in the final model. This model building approach is different from stepwise regression. The variables included were selected because of the theoretical considerations and were only dropped from the model if they caused statistical problems with one or more other variables in the model.
Test for Multicollinearity

Variance inflation factor (VIF) and tolerance tests in statistical software SAS, indicated higher degree of multicollinearity among pine sawtimber, pine pulpwood, hardwood sawtimber and hardwood pulpwood prices. In the SAS output VIF value greater than 10 prove multicollinearity. A tolerance value lower than 0.1 is comparable to a VIF of 10. Pine sawtimber and pine pulpwood prices are more important for studying the harvesting behavior from the theoretical point of view. Therefore other three product prices were dropped from the independent variables. But there was still significant multicollinearity present among the pine sawtimber and pine pulpwood prices. Hence pine pulpwood price variable was dropped because sawtimber prices are more important as the main commercial product from the final harvest is pine sawtimber. The other variables were retained in order of their theoretical and statistical importance (Pseudo $R^2$ from bivariate model). The final variables retained in the multivariate logistic regression were sawtimber prices, softwood volume, softwood proportion, number of large trees, proportion of large trees, slope, categorical variables for land productivity, hydric region, piedmont region, economic ranking and natural pine (bmt=2).

The Model Convergence Status in the SAS results suggested that the maximum-likelihood algorithm converged. Three asymptotically equivalent Chi-Square tests of likelihood Ratio, score and wald test were estimated. The above mentioned tests were used to test the null hypothesis that at least one of the predictors’ regression coefficient is not equal to zero in the model. All these tests rejected the null hypothesis. The pseudo $R^2$ for the model was 0.142. Comparing this to the previous studies for the same region it was slightly better
(Polyakov et al 2010 had Pseudo $R^2$ of 0.12 for North Carolina). The higher explanatory power was likely due to the fact that the price data used in this analysis was more heterogeneous (7 micro markets in North Carolina) as compared to the price data used in the earlier studies (2 regions in North Carolina). Also, product prices were used as one of the predictors instead of revenue. Based upon the timber supply theory some additional predictors such as number and proportion of large trees, housing prices, economic rank of the county, hydric soil, dummy variable for piedmont region and land productivity were used in the model.

**Results for Multivariate Model**

The results from multivariate logistic model suggested that softwood volume, softwood proportion, pine sawtimber prices, number and proportion of large trees all have positive impact on the probability of harvest. The variables for piedmont region, economic tier, productivity, slope and housing prices had expected signs but were statistically insignificant (as the confidence interval associated with their estimate included 1). The results (table 6.) proved that with every unit increase in soft wood volume and saw timber prices the probability of harvest increases by 3% (odds ratio=1.03) and 7.3% (odds ratio=1.073) respectively. For the softwood proportion variable, each unit increase would result in a 100.8% (odds ratio=2.008) increase in harvest probability. Similarly with each unit increase in large tree proportion variable the harvest probability increases 6.49 times or 549% (odds ratio=6.49) and with unit increase in large tree number the harvest probability increases by 1.005 times (odds ratio=1.005).
The predictor variables that were negatively correlated with the probability of harvest were dummy variables for hydric region and natural pine. The probability of harvest of a plot in hydric class was 60.6% (odds ratio=0.394) lesser than a plot in mesic class (base case). Similarly the harvesting probability of a natural pine plot was 64.8% (odds ratio=0.352) less than a planted pine plot (which was the base case).

The average probability of harvest for all plots was 0.11 in multivariate analysis. The data was for 7 years (2003-2010), therefore, the average annualized probability of harvest was 0.0157. The harvest probability (predicted) distribution of FIA plots was presented in figure 4. The map confirmed that most of the high and medium probability plots were located in coastal plains; a few in piedmont and very few in mountain region.

4.4. Results for Availability

Any public policy should be formulated based upon the actual availability of a resource instead of its apparent presence in a geographical region. The availability of a resource for a specific purpose depends upon its accessibility, alternate uses, biological cycle and economic factors. The resource in question in this study was wood or timber. Total wood inventory in a region does not equal the amount of wood that is actually available for conversion into wood products. Based upon the econometric analysis, this study recognizes the biophysical and economic variables that may limit wood availability on forests in North Carolina.

Wood can be defined as the dry tons of boles, tops, and limbs of all trees at least 1.0 in. dbh on forestland. (Bechtold, William A.; Patterson, Paul L., Editors. 2005)
Availability Discussion

The reason for discussion of available timber inventory and harvest probability in terms of the same variables is because of their interdependence. Timber inventory in the market area specified can be described as a stock (only a portion of which is actually available) and timber harvest can be described as a flow/supply from that stock. The available inventory has a positive impact on the harvest probability. This is because of the reason that more available inventory means the economies of scale and less negative impact on the amenity values. The variables which are related positively to the availability are expected to have same signs in the harvest probability model. Thinking from another way, harvest probability has a direct impact on the inventory availability.

As discussed in chapter 3, it is not possible to determine a single threshold for a variable (e.g. percentage slope) to separate available timber from unavailable timber. The threshold depends upon multivariate impacts of biophysical variables, market conditions, species in question, ownership objectives etc. For example if the market value of a species is so high that it is profitable to harvest it even at high costs (e.g. on steep slopes) then it would be available. On the other hand if the market value of a species is low then it is possible that it would not be harvested even on easily operable sites, and then it would be unavailable.

Econometric analysis and timber availability

The empirical variables and estimates for timber availability discussion were based upon the results from the logistic regression results. The availability was investigated in terms of the impact of the continuous predictors (pine saw timber prices, slope, softwood
volume, softwood proportion and large tree proportion) on harvest probabilities (figures 6-10). Graphs were constructed exhibiting the relationship between harvest probability and a predictor variable. The blue curve was derived from the bivariate models (individual impact of the predictors) and the red curve was derived from the multivariate analysis.

For deriving the multivariate curve the parameter estimates from the multivariate logistic regression results were used. The individual predictor curves were formed by keeping all the other variables at their mean value except the variable in question.

**Sawtimber prices**

Sawtimber price was one of the chief economic variables used in the econometric analysis. Both bivariate and multivariate curves in the graph (figure 6) demonstrated that when sawtimber price became less than $35/ton, harvest probability had fallen sharply. The multivariate curve was flatter in comparison to bivariate curve (based upon the existing price data range) since it accounted for other factors than price. The supply curve result with respect to prices was interesting because in the research literature the elasticity of supply is significantly lower than the cross-sectional analysis in this study exhibited. One of reasons was that this study was based upon a price data which exhibited more regional variation than most of the previous studies exist in literature (Polyakov et al. 2010, Butler 2006, Prestemon and Wear 2000). The other reason may be the absence of ownership information in this study.

**Slope**

The impact of slope on harvest probability was shown in graph (figure 7). It was evident from the bivariate curve that when the slope percentage was more than 40 the
probability of harvest became too low (approx. 0.06) compared to the mean value (0.11) in multivariate analysis. This result was based upon bivariate analysis, but in real world the impact is always multivariate (product prices, land productivity, tree species etc.). For the same probability the multivariate threshold for slope percentage was 60. The reason for higher slope threshold with multivariate estimates was due to the inclusion of other predictors such as sawtimber prices. For example if the product price is higher than it would be profitable to harvest it even on steeper slopes. The estimate range of above thresholds were agreeable with the harvesting guidelines for some states (Massachusetts and New York states stipulate harvesting restrictions on slopes above 60% and 30% respectively).

**Softwood Proportion**

The bivariate curve in the graph for softwood proportion and harvest probability (figure 8) demonstrated that harvest probability increases with increase in softwood proportion in the stand and vice versa. If the proportion of softwood went below 0.2, the harvest probability became significantly low compared to the mean value of 0.11. The multivariate curve was flatter and had a lower threshold (approx. 0.35) because of the inclusion of other biophysical, economic and location factors. The results proved the market importance of softwood stands.

**Softwood Volume**

Both bivariate and multivariate curves (figure 9) demonstrated that if the plot softwood volume per acre was below 500 cuft then there would be very low probability (as compared to the mean probability from the multivariate model which was 0.11) of it to be a
part of available inventory. The multivariate curve was flatter because of the simultaneous effect of all the predictors.

**Large tree proportion**

Large tree proportion (per acre) in a plot was used as a proxy for age distribution. The large trees are critical for the better industrial end products and cost efficient harvesting techniques. The graph demonstrated the importance of large size trees (dbh > 9 inches) for harvest probabilities. There was not much difference in the curves for bivariate and multivariate estimates (figure 10.). The curves exhibited that harvest probability for the plot would be significantly low (as compared to mean value of 0.11) if the large tree proportions fall below 0.2.

**4.5. Discussion**

Timber supply is a complex process. Long life cycles, type of ownership / preferences, environmental factors, economic factors (prices costs etc.) and general economy of the region are some of the factors from a long list which make it multidimensional. The current study can be classified as a positive harvest model (Wear 1994) based upon the individual harvest behavior. It was similar in basic methodology to the earlier attempts by various researchers for example Prestemon and Wear (2000) and Polyakov et al. (2010) etc. On the other hand it was different from the above mentioned studies in terms of current updated inventory estimates, use of variables related to opportunity cost (economic tier and housing prices), variables related to age structure (proportion of large trees), better price data
(explained in the methodology section) for econometric (cross-sectional) analysis, attempt to estimate availability in scientific manner and this study was exclusively for North Carolina State.

Opportunity cost is a very important aspect which effects timber supply. In this study the impact of opportunity cost was investigated by including economic tier/status of the county and the average housing prices in the county. Construction of houses and urbanization represent the alternate use of forest land. The results suggested that hypothesis for negative impact of higher housing prices and more urbanization was correct (as both these variables represented opportunity cost or alternate use of forest land). In other words if the income or value of land in other uses is more, the forest land/timber supply/harvest would be low.

The other economic, physiographic and location variables had expected signs. Softwood volume (SV) and softwood proportion (Sprop) variables had positive correlation with harvest probability which proved that softwoods is economically more important and has wider markets than hardwoods. The results related to the impact of physiographic class suggested that plots at hydric sites have very low probability of harvest. The negative impact of slope on harvesting probability proved that as the percentage slope increases timber supply decreases because of the difficulty in planting and harvesting operations. The plots classified as low productivity sites had negative impact on harvesting probability and hence timber supply. For the categorical variables of management classes, all the management types were compared with the planted pine (bmt=1). In the bivariate model the management types were compared individually with the planted pine and found to have lower probability of harvest. But differences with upland and lowland hardwoods were not statistically significant.
Therefore based upon the significance and pseudo $R^2$ for bivariate model only natural pine (bmt=2) was included in the multivariate model which was found to be negatively correlated with the probability of harvest.

Among the financial/economic variables pine sawtimber, pine pulpwood and chip-n-saw prices were statistically significant and positively correlated with the harvest probability in bivariate model. This result was consistent with timber supply theory as softwood is more important product from market point of view than hardwood. These three prices were found to be collinear in the multivariate analysis, hence only saw timber price was chosen (based upon the coefficient value and bivariate pseudo $R^2$) in the final multivariate model.

The impact of location (physiographic region) on harvest probability was examined by including categorical variables representing northern coastal plain region, piedmont region, mountain region and southern coastal plain region (Intercept or base case). The multivariate analysis results suggested that probability of harvest in southern coastal plain was higher than piedmont and mountain regions. It was lowest in mountains.

The limitation for this study was the unavailability of detailed ownership data. Ownership is a very important variable to examine harvest behavior. The ownership type is important because the objectives for the non-industrial owners and industrial owners differ in a sense that timber harvesting decisions also depends upon non timber amenities (for example recreation, hunting) and these non-timber amenities are correlated with the structure of forest capital and ownership (Pattanayak, Murray and Abt 2001).

The econometric analysis of harvesting probabilities had important consequences for the estimation of timber availability part of the study. The variables used to discuss the
timber availability in North Carolina were selected on the basis of econometric results. The harvest probability graphs for continuous variables (sawtimber prices, slope, softwood volume, softwood proportion and large tree proportion) suggested rough thresholds to decide what portion of inventory is available and what is non-available. But it was hard to choose a clear single threshold. This was because of the reason that even if a predictor (steep slope) points towards non availability another predictor (very high prices) might contradict it.
CHAPTER 5 – CONCLUSIONS

The empirical results in this study exhibited that biophysical (soft wood volume, softwood proportion, large tree proportion and number, land productivity, hydric region, physiographic region) and economic (opportunity cost and prices) factors are important predictors for harvest behavior. The important conclusions from this study were:

- Forest inventory data supplemented with other economic data (based upon the economic theory), for example economic status of the county where the plot is located can result into important insights about the timber supply process.

- Stand age structure is a critical predictor of timber supply, as the larger trees have more market value. Also, higher proportion of large trees in the plot makes harvest cost efficient. The empirical result for the large tree proportion variable (used as a proxy for age structure) proved the above hypothesis.

- The price data used in this study had more regional variation as compared to the data used by researchers in the past. More regional variation was the primary reason behind the significantly large cross-sectional supply elasticity (figure 6).

- Product prices were directly used as predictors for harvest probability. Whereas similar cross-sectional studies in past used revenue function as predictor. The issue with using revenue (price*volume) instead of prices is that it represents the joint impact of volume and prices, therefore by treating both coefficients as equal. As this study demonstrated they are not equal and should be examined separately.
• Hardwood volume was proved to be a non-significant predictor for harvest probability. Hardwood prices were also found to be negatively correlated with the harvest probability. These two results were surprising because hardwoods constitute a major portion of overall timber harvest in North Carolina. Therefore hardwoods should be studied as a separate product in terms of harvesting behavior.

• Use of additional variables derived from FIA data (large tree proportion and number) supplemented with better price data, use of price as separate predictor from volume and use of proxy variables for opportunity cost resulted into higher pseudo $R^2$ (0.142) than similar studies in past.

• Timber availability thresholds were discussed based upon the empirical estimates rather than using the estimates from past studies or expert opinions.

• There were data constraints in this study, for example, non-availability of ownership data.

The discussion in this study illustrated that it is possible to discuss timber availability from a region based upon the econometric results, but it is difficult to formulate a single threshold for availability because of the multidimensionality and complexity of the supply function. The study can be improved by including ownership characteristics (for example type, size, education etc.), and inspecting the structural shift in the markets and timber supply due to the lower prices of the economic recession years (2006-09).
REFERENCES


BINKLEY, C.S. 1981. Timber supply from private nonindustrial forests. Yale University, School of Forestry and Environmental Studies Bulletin 92.


HARTMAN, R. 1976. The harvest decision when the standing forest has value. Econ. Inquiry 14(1):52–58.


HUNT, L. 2002. Exploring the availability of Ontario’s non-industrial private forest


MILES, P.D. 2009. Forest Inventory EVALIDator web-application version 4.01 beta.


UNDERSTANDING FORESTRY TERMS – A GLOSSARY FOR PRIVATE LAND OWNERS. Available online at:


Table 1. Descriptions and summaries of biophysical variables (Continuous).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Exp. Sign</th>
<th>Units or Categories</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softwood Volume (SV)</td>
<td>+</td>
<td>100 cubic ft/ ac</td>
<td>5.88 (9.5)*</td>
<td>0</td>
<td>63.12</td>
</tr>
<tr>
<td>No. of Large (dbh &gt; 9 in) trees per acre</td>
<td>+</td>
<td>Number</td>
<td>60 (46.67)</td>
<td>0</td>
<td>313</td>
</tr>
<tr>
<td>Proportion of Large (dbh &gt; 9 in) trees Per acre</td>
<td>+</td>
<td>Proportion</td>
<td>0.17 (0.21)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Softwood Proportion (SProp)</td>
<td>+</td>
<td>Proportion</td>
<td>0.40 (0.40)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Slope</td>
<td>-</td>
<td>Percentage</td>
<td>10.2 (16.7)</td>
<td>0</td>
<td>143</td>
</tr>
</tbody>
</table>

*Figures in brackets under the column “mean” represent standard deviation.
Table 2. Descriptions and summaries of biophysical and location variables (Categorical).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Exp. Sign</th>
<th>Categories</th>
<th>No. of Plots(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management Type</td>
<td>Ref</td>
<td>Planted Pine</td>
<td>14.34</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Natural Pine</td>
<td>14.61</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Oak Pine</td>
<td>11.33</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Hardwood (U)</td>
<td>45.15</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Hardwood (L)</td>
<td>14.57</td>
</tr>
<tr>
<td>Physiographic Region</td>
<td>Ref</td>
<td>S. Coastal Plain</td>
<td>28.39</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>N. Coastal Plain</td>
<td>20.34</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Piedmont</td>
<td>33.79</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Mountain</td>
<td>17.49</td>
</tr>
<tr>
<td>Productivity</td>
<td>Ref</td>
<td>Low</td>
<td>12.32</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>High</td>
<td>87.6</td>
</tr>
<tr>
<td>Physiographic Class</td>
<td>+</td>
<td>Xeric</td>
<td>6.77</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Mesic</td>
<td>87.15</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Hydric</td>
<td>6.08</td>
</tr>
</tbody>
</table>
### Table 3. Descriptions and summaries of economic variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expected Sign</th>
<th>Units</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pine sawtimber price</td>
<td>+</td>
<td>$/ton</td>
<td>35.80 (3.14)*</td>
<td>28.73</td>
<td>39.29</td>
</tr>
<tr>
<td>Pine pulpwood price</td>
<td>+</td>
<td>$/ton</td>
<td>6.79 (0.90)</td>
<td>14.86</td>
<td>23.4</td>
</tr>
<tr>
<td>Chipnsaw price</td>
<td>+</td>
<td>$/ton</td>
<td>20.27 (2.44)</td>
<td>5.41</td>
<td>8.27</td>
</tr>
<tr>
<td>Hardwood Sawtimber price</td>
<td>+</td>
<td>$/ton</td>
<td>20.67 (0.66)</td>
<td>19.88</td>
<td>21.8</td>
</tr>
<tr>
<td>Hardwood Pulpwood price</td>
<td>+</td>
<td>$/ton</td>
<td>4.64 (0.92)</td>
<td>3.87</td>
<td>7.34</td>
</tr>
<tr>
<td>Housing Price</td>
<td>-</td>
<td>$ 1000</td>
<td>238.76 (85.52)</td>
<td>105.26</td>
<td>526.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expected Sign</th>
<th>Categories (Econ.Tier)</th>
<th>No. of Plots (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom</td>
<td>Ref</td>
<td>Bottom</td>
<td>43.75</td>
</tr>
<tr>
<td>Top</td>
<td>-</td>
<td>Top</td>
<td>56.24</td>
</tr>
</tbody>
</table>

*Figures in brackets under column “Mean” represent standard deviation.*
Table 4. Bivariate logistic regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
<th>Pseudo R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softwood Volume</td>
<td>1.066</td>
<td>(1.056 1.077)*****</td>
<td>0.079</td>
</tr>
<tr>
<td>Softwood Proportion</td>
<td>3.614</td>
<td>(2.744 4.759)*****</td>
<td>0.041</td>
</tr>
<tr>
<td>Hardwood Volume</td>
<td>0.989</td>
<td>(0.979 0.998 )</td>
<td>0.00001</td>
</tr>
<tr>
<td>PineSaw Price</td>
<td>1.139</td>
<td>(1.087 1.194)*****</td>
<td>0.017</td>
</tr>
<tr>
<td>Pulpwood Price</td>
<td>1.356</td>
<td>(1.202 1.529)*****</td>
<td>0.011</td>
</tr>
<tr>
<td>ChipnSaw Price</td>
<td>1.161</td>
<td>(1.096 1.230)*****</td>
<td>0.014</td>
</tr>
<tr>
<td>HardSaw Price</td>
<td>0.576</td>
<td>(0.481 0.690)*****</td>
<td>0.017</td>
</tr>
<tr>
<td>HardPulp Price</td>
<td>0.754</td>
<td>(0.655 0.868)*****</td>
<td>0.008</td>
</tr>
<tr>
<td>Stumpage Value</td>
<td>1.013</td>
<td>(1.007 1.019)*****</td>
<td>0.01</td>
</tr>
<tr>
<td>Top tier</td>
<td>0.742</td>
<td>(0.596 0.923)*****</td>
<td>0.003</td>
</tr>
<tr>
<td>House Price</td>
<td>0.997</td>
<td>(0.996 0.999)*****</td>
<td>0.006</td>
</tr>
<tr>
<td>Mountain</td>
<td>0.314</td>
<td>(0.207 0.477)*****</td>
<td>0.0252</td>
</tr>
<tr>
<td>Slope</td>
<td>0.974</td>
<td>(0.965 0.984)*****</td>
<td>0.018</td>
</tr>
<tr>
<td>N. Coastal Plain</td>
<td>1.32</td>
<td>(1.021 1.705)**</td>
<td>0.002</td>
</tr>
<tr>
<td>Piedmont</td>
<td>0.733</td>
<td>(0.576 0.934)*****</td>
<td>0.003</td>
</tr>
<tr>
<td>Hydric</td>
<td>0.401</td>
<td>(0.210 0.767)*****</td>
<td>0.004</td>
</tr>
<tr>
<td>Productivity</td>
<td>2.646</td>
<td>(1.663 4.21)*****</td>
<td>0.01</td>
</tr>
<tr>
<td>bmt2</td>
<td>0.687</td>
<td>(0.486 0.970)*****</td>
<td>0.002</td>
</tr>
<tr>
<td>bmt3</td>
<td>0.697</td>
<td>(0.474 1.025)**</td>
<td>0.0023</td>
</tr>
<tr>
<td>bmt4</td>
<td>0.962</td>
<td>(0.772 1.198)</td>
<td>0.0001</td>
</tr>
<tr>
<td>bmt5</td>
<td>0.777</td>
<td>(0.557 1.082)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Large tree prop.</td>
<td>3.56</td>
<td>(4.006 8.285)*****</td>
<td>0.025</td>
</tr>
<tr>
<td>No. of large trees</td>
<td>1.009</td>
<td>(1.006 1.009)*****</td>
<td>0.044</td>
</tr>
</tbody>
</table>

***p<0.01; **p<0.05; *p<0.10; Odds ratio >1 means positive sign; odds ratio < 1 means negative sign.
Table 5. Multivariate logistic regression (complete set of predictors).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
<th>Pseudo R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softwood Volume</td>
<td>1.007</td>
<td>(1.000 1.0013)**</td>
<td>0.166</td>
</tr>
<tr>
<td>Softwood Proportion</td>
<td>4.2</td>
<td>(2.41 7.32)**</td>
<td></td>
</tr>
<tr>
<td>Hardwood Volume</td>
<td>1.001</td>
<td>(0.996 1.007)</td>
<td></td>
</tr>
<tr>
<td>PineSaw Price</td>
<td>1.024</td>
<td>(0.577 1.819)</td>
<td></td>
</tr>
<tr>
<td>Pulpwood Price</td>
<td>1.383</td>
<td>(0.747 2.560)</td>
<td></td>
</tr>
<tr>
<td>ChipnSaw Price</td>
<td>0.928</td>
<td>(0.501 1.721)</td>
<td></td>
</tr>
<tr>
<td>HardSaw Price</td>
<td>1.77</td>
<td>(0.574 5.461)</td>
<td></td>
</tr>
<tr>
<td>HardPulp Price</td>
<td>0.75</td>
<td>(0.479 1.175)</td>
<td></td>
</tr>
<tr>
<td>Stumpage value</td>
<td>1.004</td>
<td>(0.995 1.013)</td>
<td></td>
</tr>
<tr>
<td>Top tier</td>
<td>0.743</td>
<td>(0.417 1.324)</td>
<td></td>
</tr>
<tr>
<td>House price</td>
<td>0.999</td>
<td>(0.997 1.001)</td>
<td></td>
</tr>
<tr>
<td>Mountain</td>
<td>0.201</td>
<td>(0.079 0.512)**</td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.991</td>
<td>(0.977 1.006)</td>
<td></td>
</tr>
<tr>
<td>N. Coastal Plain</td>
<td>0.491</td>
<td>(0.259 0.931)**</td>
<td></td>
</tr>
<tr>
<td>Piedmont</td>
<td>0.403</td>
<td>(0.227 0.718)**</td>
<td></td>
</tr>
<tr>
<td>Hydric</td>
<td>0.397</td>
<td>(0.194 0.814)**</td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td>1.715</td>
<td>(1.058 2.910)**</td>
<td></td>
</tr>
<tr>
<td>Bmt2</td>
<td>0.471</td>
<td>(0.300 0.738)**</td>
<td></td>
</tr>
<tr>
<td>Bmt3</td>
<td>0.904</td>
<td>(0.544 1.503)</td>
<td></td>
</tr>
<tr>
<td>Bmt4</td>
<td>2.633</td>
<td>(1.662 4.171)**</td>
<td></td>
</tr>
<tr>
<td>Bmt5</td>
<td>1.304</td>
<td>(0.752 2.662)</td>
<td></td>
</tr>
<tr>
<td>Large tree prop.</td>
<td>6.12</td>
<td>(3.595 10.41)**</td>
<td></td>
</tr>
<tr>
<td>No. of large trees</td>
<td>1.009</td>
<td>(1.006 1.012)**</td>
<td></td>
</tr>
</tbody>
</table>

***p<0.01; **p<0.05; *p<0.10; Odds ratio >1 means positive sign; odds ratio < 1 means negative sign.
Table 6. Multivariate logistic regression (Final).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
<th>Pseudo R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Softwood Volume</td>
<td>1.03</td>
<td>(1.011  1.049)***</td>
<td>0.142</td>
</tr>
<tr>
<td>Softwood Proportion</td>
<td>2.008</td>
<td>(1.252  3.219)***</td>
<td></td>
</tr>
<tr>
<td>PineSaw Price</td>
<td>1.073</td>
<td>(1.006  1.144)**</td>
<td></td>
</tr>
<tr>
<td>Top tier</td>
<td>0.87</td>
<td>(0.680  1.114)</td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.987</td>
<td>(0.97  1.000)**</td>
<td></td>
</tr>
<tr>
<td>Piedmont</td>
<td>0.775</td>
<td>(0.590  1.017)**</td>
<td></td>
</tr>
<tr>
<td>Hydric</td>
<td>0.394</td>
<td>(0.198  0.783)***</td>
<td></td>
</tr>
<tr>
<td>Productivity</td>
<td>1.151</td>
<td>(0.947  2.539)**</td>
<td></td>
</tr>
<tr>
<td>bmt2</td>
<td>0.352</td>
<td>(0.238  0.521)**</td>
<td></td>
</tr>
<tr>
<td>Large tree prop.</td>
<td>6.49</td>
<td>(3.86  10.91)***</td>
<td></td>
</tr>
<tr>
<td>No. of large trees</td>
<td>1.005</td>
<td>(1.002  1.009)***</td>
<td></td>
</tr>
</tbody>
</table>

***p<0.01; **p<0.05; *p<0.10; Odds ratio >1 means positive sign; odds ratio < 1 means negative sign.
Figure 1. Distribution of FIA plots in North Carolina.
Figure 2. Distribution (mean, median, interquartile range) of harvesting intensities (excludes plots with no trees harvested)
Figure 3. North Carolina Survey units (Physiographic units)
Figure 4. Predicted harvest distribution of plots in North Carolina (High = 99% quantile; Med = 90% quantile; Low = 75% quantile)
Figure 5. Actual harvest distribution of planted pine plots in North Carolina.
Figure 6. Sawtimber prices versus harvest probability
Figure 7. Slope versus harvest probability.
Figure 8. Softwood proportion versus harvest probability.
Figure 9. Softwood volume versus harvest probability.
Figure 10. Proportion of large trees versus harvest probability.