2007

Factors influencing the spatial distribution of natural resource-based industries: the softwood lumber industry in the United States South

Francisco Xavier Aguilar
Louisiana State University and Agricultural and Mechanical College

FACTORS INFLUENCING THE SPATIAL DISTRIBUTION OF NATURAL RESOURCE-BASED INDUSTRIES: THE CASE OF THE SOFTWOOD LUMBER INDUSTRY IN THE UNITED STATES SOUTH

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

In

The School of Renewable Natural Resources

by
Francisco X. Aguilar
B.S., Universidad Escuela de Agricultura de la Region Tropical Humeda, 1998
M.S., Royal Agricultural College, 2002
M.S., Louisiana State University, 2006
May 2007

ACKNOWLEDGMENTS

I would like to thank all members of my Doctoral Committee for their guidance and mentoring. I found their advice along the entire process that involves a Doctoral Program essential to any success I may have accomplished. I am particularly grateful to Professor Richard Paul Vlosky for his support to my research, his mentoring, and for giving me the freedom and financial support to pursue my research interests and career goals.

All my previous experiences have been crucial to crafting my career and abilities as a scholar but I place particular importance on my time working with Resources for the Future (RFF) in Washington D.C. I am indebted to Dr. Allen Blackman, Dr. Carolyn Fisher, and Dr. Roger Sedjo for their advice and for giving me the opportunity to be a colleague. I am also very thankful to RFF for granting me a J.L. Fisher Dissertation Fellowship. The Fellowship partly funded my research and represents a startling point in my academic career.

Thanks to Dr. Richard Kazmierczak Jr. for his words of advice and constant support. I would also like to thank Lisa Mauger and Borja Servan for their help figuring out some of my statistical analyses. Their minds are definitely brighter than mine sorting out codes!

Tremendous thanks to my "support" team. I have to start with Dr. Denese Ashbaugh-Vlosky who recruited me to come to LSU and who is a mother-like figure to me. Virginia and Charles Grenier have become some of my dearest friends and helped keeping me up during down times. Thanks go to my officemates Shadia, Sanna, and Ben for their friendship and encouragement. Also, thanks to Mau vanDuren and Jackie Coolidge, my host family and great friends in D.C., and the Purgersons in Baton Rouge. And last but never least, thanks to my family. My mother, father, grandparents, brothers, and my extended family, they are an inspiration and my day-to-day motivation to excel in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Expanding on the Theory of Location, New Economic Geography, and Porter's Theory of Clusters this dissertation attempts to identify key factors influencing the location of firms in a resource-based industry. The softwood lumber industry in the United States is used as a case study to test several hypotheses concerning these theories. Two U.S. regions, the West and the South, were selected for analysis because they account for over 70 percent of U.S lumber manufacturing capacity.

A multi-disciplinary research approach involved three-stages. First, self-reported preferences were analyzed using common factor and conjoint analyses for preferences for location attributes. Surveys were sent to all sawmill managers in the U.S. West and South regions. Respondents were identified through the Random Lengths' Big Book (2006), the industry's most comprehensive database. Survey procedures followed Dillman's (2000) Tailored Design Method. Conjoint analysis provided information on the relative importance of selected site attributes using several econometric models to estimate coefficients, significance and marginal effects of site attributes.

Second, a model for industry location behavior in the U.S. South was developed using a spatial econometric model. An exploratory analysis identified deviations from complete spatial randomness as first-hand evidence of clustering. The presence of sawmill enterprises was used as the dependent variable, aggregated at the county level. Spatial autoregressive and correlated error econometric models were used to correct for spatial correlation. The final model was used to identify counties where softwood lumber industry development could occur in the future with a high probability of success.

Third, two cross-sections of data were analyzed using point density tools to explore spatial concentration in the softwood lumber industry over time. There is evidence of

consolidation in the industry as the number of firms has declined while capacity has increased over time.

The findings are congruent with spatial predictions drawn by Location Theory, New Economic Geography and to some extent the Theory of Clusters. Research methods used in this study have the ability to capture decision-makers preferences and to operationalize major theories of location, economic geography and cluster development. Results can provide industry and economic development professionals with a new decision-making tool.

CHAPTER 1. INTRODUCTION

The study of the factors influencing the geographic location of industries has received considerable attention from the scientific community. Edward Ross's (1896) seminal paper published in 1896 is probably one of the first to present a theory to explain the "Location of Industries". Ross suggested that enterprises locate in particular sites because of economic, rational, and non-rational reasons. Some non-rational reasons may be simply accident or caprice, but firms mainly choose a location over others because it is deemed to offer particular economic and rational benefits, such as lower production costs or marketing opportunities. Ross (1896) further stresses that entire industries, compared to individual enterprises, choose a location based on specific economic advantages and not on non-rational or personal causes.

Baker (1926) considered that the study of the economics of manufacturing demands the formal analysis of the physical conditions and location of land. Geographic knowledge provides an understanding of the origin and type of materials used as well as the relative advantages and disadvantages of different locations for a factory, both with reference to purchase of inputs and sale of finished goods. Weber's (1929) theory of location of manufacturing industries suggests that the factors determining the location of manufacturing enterprises are conceived of as being specific cost advantages in certain places. Weber based his theory of location of industries on the concept of minimum transportation costs (Kennelly 1968). Conditions affecting transportation costs include weight of goods, existing technology, and distances to markets, road conditions and the nature of the good that is transported. Weber considers general location factors to be the cost of grounds, cost of buildings, machines and other fixed capital costs, costs of securing materials, power and fuel, cost of labor, cost of transportation, interest rates and rate of depreciation on fixed capital (Kennelly 1968). Predohl (1928) groups all types of costs in three classes which are cost of transportation, cost of raw material and cost of labor. The causes that affect the location

of an enterprise and the spatial distribution of an industry are diverse and vary depending on the type of industry. According to Renner (1931) economic advantages at a particular location vary depending on whether the nature of an industry is extractive (i.e. mining, logging), reproductive (i.e. agriculture, forestry), fabricative (i.e. manufacturing) or facilitative (i.e. service industries).

Ross (1896) made a distinction between forces affecting location and industry localization. Location refers to the decision of an industry to decide to be in a particular place because of advantages derived from access to natural resources or the presence of input suppliers. Localization instead is not specific to location and refers to the economics gains from industry clustering that can emerge as a result of interaction between enterprises. Alfred Marshall (1920) identified three main forces driving industrial agglomeration (localization). The first is the presence of a large labor market pool. The second component is the provision of intermediate goods and services. These include raw materials, supplies, consultations and collaboration. The last component is the occurrence of knowledge exchanges and spillovers between nearby firms and institutions. More recently, Ellison and Glaeser (1997) provided empirical evidence of localization across manufacturing industries, they consider that industrial geographic concentration is ubiquitous and the most extreme cases of localization are likely due to pure natural advantages.

A better understanding of the causes behind industry localization and factors influencing spatial distribution of resource-based industries can support the planning process for an industry and help in crafting regional development policies. Hoover (1948) considers that the intentional influencing of location may serve various objectives including (1) to increase the total productive capacity or total income of the area of interest, (2) to generate a more desirable combination of economic activities in the area such as the selective encouragement of new industries that can employ previously unemployed labor, and (3) to improve the processes of locational selection

and adjustment to locational change by providing information about sites and job opportunities. Although the actual decision as to where to locate a new enterprise lays in the hands of business entrepreneurs, public policy can influence their decisions indirectly. The factors of prospective profits that influence an executive's decision are inevitably affected by the actions of public authorities, whether or not these proceed upon the basis of any clearly defined locational policy (Hoover 1948). It is this indirect determination of the conditions under which private businesses make their decisions that constitutes the chief sphere of public locational policy.

The ultimate goals of locational policy are the full and continuous use of production factors, good all-round living and working conditions; individual economic security; variety of individual economic opportunity; national solidarity, security and power; and steady economic progress (Hoover 1948). Hamilton (1971) considers that location policy aims to promote national security, improve socio-economic conditions of minority groups, aid backward regions, achieve an even industrial dispersion throughout the national territory, reduce socio-economic gaps between urban and rural areas, locate resource processing near their production sources, serve market areas from central places, and achieve regional specialization and self-sufficiency.

This doctoral dissertation attempts to determine the factors affecting the spatial distribution and factor preferences for the location of resource-based industries using the softwood lumber industry as an example. This industry has been chosen because of reported evidence of industrial clustering, its importance to regional economies, and its direct link to a natural resource: wood.

1.1 The Wood Products and Lumber Industry in the U.S.

In 2005, the Wood Products Industry, North American Industry Classification System (NAICS) code 321 , provided more than 539,103 jobs nationwide of which 431,113 were directly involved in the wood manufacturing process (U.S. Census Bureau 2006b). Total wages

for production workers accounted for over 12.6 billion dollars. Value added for the wood products sector accounted for 44.7 billion dollars and total costs of materials was 67.7 billion dollars. The total value of shipments for the wood products sector was over 112 billion dollars as reported by the U.S. Census Bureau (2006b). Table 1 presents data from 2002 to 2005.

Table 1. Total number of employees, production workers, wages, value added, total cost of materials and value of shipment for NAICS code 321: Wood products manufacturing.

	All employees	Production workers					
Year	Number	Number	Hours (1,000)	Wages ($1,000)	Value added ($1,000)	Total costs of materials	Total value of shipments ($1,000)
2005	539,103	431,113	911,506	12,620,931	44,762,514	67,739,453	112,017,533
2004	535,996	429,752	908,834	12,174,518	43,733,529	60,904,957	104,135,194
2003	513,900	421,725	862,269	11,349,849	37,077,225	55,032,506	92,068,903
2002	539,981	441,738	876,476	11,554,993	35,121,393	53,984,547	89,019,024

Source: U.S. Census Bureau (2006b).

Total lumber production in the United States amounted to 50.7 billion board feet in 2005 (U.S. Census Bureau 2006a). This figure represents a 2.2 percent growth compared to the 2004 lumber production of 49.6 billion board feet. The majority of lumber produced in the U.S. is softwood lumber accounting for 78 percent of total production. The remaining 22 percent is hardwood lumber. Table 2 summarizes data from 1999 to 2005 and presents the breakdown between softwood and hardwood lumber.

The two most important regions to the softwood lumber industry in the U.S. are the Southern and Western regions. As classified by the Forest Service (2005), the Southern region includes Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, and Virginia. The Western region includes Arizona, California, Colorado, Idaho, Montana, New Mexico, Oregon, South Dakota, Utah, and Washington.

The Southern region produced a total 22.4 billion board feet of total lumber representing more than 44 percent of the total national production. The Western region contributed 18.8 billion board feet of lumber, which is 37 percent of the U.S. total lumber production. Texas is classified by the U.S. Forest Service as an independent region. Texas produced 1.8 billion board feet (3.7 percent) of lumber in 2005. Data on softwood and hardwood lumber production for each state in the previously mentioned regions in 2005 is presented in Table 3.

Table 2. U.S. Lumber production from 1999 to 2005 in million board feet.

Year	Total	Softwoods	Percent of Total	Hardwoods	Percent of Total
2005	50,725	39,564	78.0	11,161	22.0
2004	49,611	38,552	77.7	11,059	22.3
2003	47,181	36,687	77.8	10,494	22.2
2002	47,499	36,377	76.6	11,122	23.4
2001	46,588	35,479	76.2	11,109	23.8
2000	49,445	37,147	75.1	12,298	24.9
1999	50,556	38,033	75.2	12,523	24.8

Source: U.S. Census Bureau (2006a).

Regarding the breakdown of softwoods and hardwoods, Southern states produced 43.5 percent of hardwood and 46.5 percent of softwood lumber in the nation (U.S. Census Bureau 2006a). Softwood and hardwood lumber production from the Western Region accounted for 47.2 percent and 3.59 percent, respectively. The latest figure of hardwood lumber production in the West should be taken with caution as most information is not disclosed by the U.S. Department of Commerce to avoid revealing data for individual companies. For details, see Table 3.

In the U.S. South, timber and agriculture, along with subsequent processing activities, contributed directly with approximately 6.0 percent of jobs and gross regional product in 1997 (Lee et al. 2002). The Wood Products Industry contributed 1.9 percent of jobs, and the agricultural sector contributed 4.3 percent of jobs in the region. Wood products accounted for 2.3 percent of Gross Regional Product, and agriculture 3.5 percent in the U.S. South in 1997.

The U.S. Wood Products Industry concentrates in the South, which accounts for 39.3 percent of all jobs associated with the industry in the U.S. Both lumber/wood products and pulp/paper spatial concentration have increased since 1969, while the furniture sector localization in this region has decreased (Lee et al. 2002).

Table 3. Lumber production of softwoods and hardwoods by state for 2004.

Region	Total lumber	Percentage	Softwood	Percentage	Hardwood	Percentage
U.S.	**49,611**	**100.00**	**38,552**	**100.00**	**11,059**	**100.00**
South	**21,767**	**43.88**	**15,541**	**40.31**	**5,158**	**46.64**
AL	2,708	5.46	2,426	6.29	282	2.55
AR	3,080	6.21	2,422	6.28	658	5.95
FL	1,068	2.15	(D)	(D)	(D)	(D)
GA	3,063	6.17	2,662	6.90	401	3.63
KY	663	1.34	14	0.04	649	5.87
LA	1,520	3.06	1,302	3.38	218	1.97
MS	2,742	5.53	2,252	5.84	490	4.43
NC	2,630	5.30	1,968	5.10	662	5.99
OK	350	0.71	333	0.86	17	0.15
SC	1,572	3.17	1,424	3.69	148	1.34
TN	889	1.79	36	0.09	853	7.71
VA	1,482	2.99	702	1.82	780	7.05
Texas	**1,792**	**3.61**	**1,568**	**4.07**	**224**	**2.03**
West	**18,501**	**37.29**	**1,336**	**3.47**	**0**	**0.00**
AZ	65	0.13	65	0.17	-	-
CA	2,962	5.97	(D)	(D)	(D)	(D)
CO	135	0.27	(D)	(D)	(D)	(D)
ID	1,694	3.41	(D)	(D)	(D)	(D)
MT	1,106	2.23	1,106	2.87	-	-
NM	(D)	(D)	(D)	(D)	-	-
OR	7,081	14.27	(D)	(D)	(D)	(D)
SD	(D)	(D)	(D)	(D)	-	-
UT	57	0.11	(D)	(D)	(D)	(D)
WA	5,236	10.55	(D)	(D)	(D)	(D)
WY	165	0.33	165	0.43	-	-

D= Information withheld to avoid disclosing data for individual companies.
Source: U.S. Census Bureau (2006a).

Forests in the U.S. South produce a variety of hardwood and softwood timber products. Softwood products dominate production with 69 percent of harvest output in 2001 (Wear et al.

2005). Among product classes, sawlogs and pulpwood products account for 41 and 42 percent of total harvest, respectively. Softwood sawlogs are the largest product class (30 percent), followed by softwood pulpwood (27 percent) and hardwood pulpwood (15 percent). These three product classes represented approximately 72 percent of total harvests in 2001 and have accounted for at least 68 percent of harvests since the 1970's (Wear et al. 2005).

The U.S. South has become a major supplier of wood and wood products for the nation; it masses over one third of the productive forestlands in the U.S. (Stokes 1997). The U.S. South produces approximately 60 percent of the nation's timber products, most of the timber sourced from private forests. The region has demonstrated strong comparative advantage in producing a renewable timber resource, as management has shifted from mining of volunteer second growth forests to intensively managed forest plantations (Wear et al. 2005). The South produces more timber than any other single country in the world, and it is projected to remain the dominant producing region in the U.S. in the future (Prestemon and Abt 2002). Total timberland area is projected to increase in many parts of the U.S. South, especially in western and northern portions, due to agricultural land conversion to forests and to tree planting. However, timberland will also be lost, especially to urban and residential land uses particularly in the Piedmont region -Virginia to Georgia- and in Florida (Prestemon and Abt 2002). Wear (2002) predicts that 10 million acres of former agricultural land will be forested between 1992 and 2020 due to future increased timber prices. As much as 25 million acres of agricultural land could be forested by the year 2040. Wear et al. (2005) foresee that any expansion in timber production in the country is expected to occur in the Southern Region.

1.2 The Softwood Lumber Industry in the U.S. Pacific Northwest and South

U.S. softwood lumber production is concentrated in the Pacific Northwest and Southern Regions. Each region has accounted for roughly one-third of national softwood output over the

last two decades (Murray and Wear 1998). The remaining third is spread throughout the rest of the country, especially in the Rocky Mountains and the Northeast. For much of the twentieth century the Pacific Northwest has been the leading center of forest products manufacturing but the adoption of cut-and-run practices that characterized earlier logging frontiers and unfavorable market conditions switched this concentration to other regions (Robbins 1985).

As pointed out previously, contrary to the Pacific Northwest, where the majority of forestlands are publicly owned, private landowners dominate the forest landscape in the South accounting for approximately 90 percent of all forestlands. Another major difference between the two major softwood producing regions in the U.S. is the type of forest that provides timber to the market. In the Pacific Northwest harvests have historically come from old growth forests (Robbins 1985), while in the South much of the harvest is derived from agricultural forestry with commercial timber rotation of 25 to 30 years (Murray and Wear 1998).

In terms of products, the softwood lumber industry in North America can be classified into three principal categories (Spelter and Alderman 2003). The largest category accounting for two-thirds of the number of mills in the industry is dimension lumber. Dimension lumber is made up of mills primarily producing nominal 2-inches (standard 38 mm) thick lumber used in light framing. The second largest category is studs. This is a subcategory of dimension lumber consisting chiefly of 2 by 3, 2 by 4, and 2 by 6 inches lumber in lengths of 7 to 10 feet (2 to 3 meters). Stud mills typically use the smallest, lowest grade logs suitable for lumber. The regional distribution is especially noteworthy: almost 40% of the volume of stud production is located in eastern Canada (Ontario, Quebec, Maritime Provinces), where it constitutes about one-third of capacity. This concentration reflects the lower quality and smaller size of the available timber in this region (Spelter and Alderman 2003). The third category is boards, pieces less than nominal 2 inches (standard 38 mm) thick, usually intended for remanufacturing into doors, millwork,

windows, and molding. The largest concentration of board mills occurs in the U.S. West where it comprises about 10 percent of the total lumber regional output (Spelter and Alderman 2003). The remaining mills include those that handle species such as cedar and redwood, mills that process timbers -lumber more than nominal 2 inches thick- and manufacturers of specialty items such as decking, fencing, and siding. Those mills whose main output is not dimension or stud are included in the third board group as presented in Table 4. Table 4 shows the softwood lumber sawmill capacity by product category in the U.S. in 2003.

Table 4. U.S. softwood lumber sawmill capacity by product category in 2003.

	Mill capacity (billion board feet) by product			
Region	**Dimension**	**Stud**	**Board**	**Total**
South	15.3	1.0	0.6	16.9
North	0.8	0.5	0.7	2.0
West	10.6	4.0	1.9	16.5
Total	26.7	5.5	3.2	35.4

Source: Adapted from Spelter and Alderman (2003).

The latest report from the U.S. Forest Service indicates that the softwood lumber industry in the United States and Canada consists of about 1,067 sawmills (Spelter and Alderman 2005). In 2005 these sawmills had a combined capacity of 80 billion board feet. In 2004, they employed about 99,000 people and produced 73.0 billion board feet of lumber. In the process, they consumed about 9.9 billion cubic feet of timber (Spelter and Alderman 2005).

The softwood lumber industry in the U.S. has experienced a shift in its structure over the last decade. Between 1996 and 2003, 149 mills closed permanently representing a loss of 7.5 billion board feet in total sawmill capacity (Spelter 2003). The industry was most heavily impacted in 2001 and 2002 during which about 50 percent of the capacity loss occurred. The closures over these two years were a consequence of interest rate increases in the U.S. that caused a decline in home-building activity and a mild recession in the general U.S. economy (Spelter 2003).

Mill closures representing 1.7 billion board feet occurred in 1998 and 1999. This loss represented a quarter of the total capacity loss over the 1996-2003 period. According to Spelter (2003), a favorable economic environment encouraged investment in mill capacity but when markets weakened in 1998, an oversupply situation developed as new capacity had come online and high-cost suppliers were pushed out of the market. Another factor that negatively impacted the industry, particularly in the U.S. West Coast and British Columbia, was a major recession of the Japanese economy. Japan represents a major export market for the Pacific Northwest region and a slowdown in its economy resulted in a decline in the demand for lumber (Spelter 2003).

Nevertheless, the combined capacity of the remaining mills has increased by 16 percent. Of the approximately total 1,140 mills, about 470 characterize their output as dimension lumber, accounting for 67 percent of capacity; 136 list studs as their primary output, representing 16 percent of the industry's volume; and 139 are primarily board mills, making up a little over 5 percent (Spelter and Alderman 2003). Capacity increases due to construction of new mills and upgrading of existing facilities also occurred in the 1996-2003 time period. Twenty five new mills were constructed representing an increase of 1186.6 million board feet (2.8 million cubic meters) in capacity. Also, 6.8 billion board feet (16.1 million cubic meters) were added to the total industry capacity due to upgrades of existing mills.

Spelter (2003) presents a summary of softwood sawmill capacity loss due to closures and net capacity gain using 1995 capacity as a baseline (Table 5) for the three major producing regions of the U.S. The U.S. West was the worst hit region because of withdrawals of federal timber from the market in the early 1990s. Harvest restrictions on public forest lands from legislation and litigation caused a sharp reduction of total lumber production in the Pacific Northwest. According to Braden et al. (1998) total lumber production has declined by 15.6 and 39.2 percent from 1987 to 1996 in the states of Washington and Oregon, respectively.

Lumber and panel production comprise the large majority of output from the solid wood sector in the U.S. South. These products utilize about 46 percent of fiber products in the South and the location of mills is widely dispersed compared to the pulp and paper sector that presents a more aggregated spatial distribution (Wear et al. 2005).

Table 5. Softwood sawmills capacity loss due to closures and net capacity gain as a percentage of 1995 capacity by U.S. region.

	1995-2002 capacity loss	1995 capacity	1995-2002 capacity loss	2002 Capacity	1995-2002 net capacity gain
	(Billion board feet)		(%)	(Billion board feet)	(%)
U.S. West	3.7	17.4	21.3	17.5	0.5
U.S. South	1.4	16.3	8.6	18.5	13.5
U.S. North	0.3	1.8	14.0	1.9	7.1

Source: Adapted from Spelter (2003).

The U.S. South was the region with the highest capacity growth due to technology advances and transfer of capacity from the U.S. West (Spelter 2003). Unlike southern pulpwood capacity, softwood sawmill capacity has not declined in the U.S. South over the last decade. Softwood sawmill capacity remained stable to slightly increasing since 2000, even as capacity in the U.S. West declined.

1.3 Geographic Concentration in the U.S. Wood Products Industry

The Wood Products Industry in the U.S. is considered to have cluster characteristics. Porter (2003) ranks the Wood Products Industry among the top 25 largest clusters in the country based on the number of people employed and spatial concentrations. Braden et al. (1998) and Porter (1998a) provide anecdotal evidence of clustering in the wood products industry in the Pacific Northwest and North Carolina, respectively.

Braden et al. (1998) identify various wood product manufacturing clusters in the U.S. Pacific Northwest. These include the log home industry in western Montana, molding and millwork industry in Oregon, and the wooden boat building industry on the Olympic Peninsula of Washington. Braden et al. (1998) consider that these clusters have emerged in locations near

and with easy access to major regional markets. For example, the molding and millwork industry in Bend, Oregon, first developed as a group of mills located close to railroad lines. The log home industry in western Montana is located along interstate highways and in a region where log homes are considered fashionable. A second characteristic necessary for the development of wood manufacturing clusters is a location where supply of raw materials and potential consumers are plentiful.

Another necessary condition for the emergence of wood product manufacturing clusters is the presence of skilled labor. Braden et al. (1998) indicate that skilled workers already resided in the areas where the clusters in Montana, Oregon and Washington developed. Clusters continued to attract skilled workers to the area as they grew larger, which in turn fostered cluster thrive. Another common factor among the clusters studied by Braden et al. (1998) is the presence of energetic entrepreneurship and mutual cooperation. Innovation is another key factor to the successful development of a cluster. Innovation helps maintain competitive advantages by developing new products. During the emergence of a new cluster competition levels seem to be low. Each cluster was the first to provide a particular good or service in their respective regions. As a cluster grows, new businesses identify niches for new or differentiated products that complement already available products.

Markets for wood inputs might best be described as localized or spatially differentiated in the Hotelling tradition because transportation costs are a large component of the delivered cost of wood (Murray 1995). Hotelling's (1929) model explains the location and pricing behavior of firms. Hotelling's model assumes that products are homogeneous and consumers buy from the least expensive location, taking transportation costs into account.

Sawmills also benefit from economies of scale, and the spatial aggregation of wood product manufacturers can result in gains in efficiencies and cost reductions (Murray 1995).

According to Cohen and Paul (2005) spatial and industrial agglomeration has external economies of scale effects in the sense that they increase, or offset, internal scale economies (Cohen and Paul 2005). In industries that experience increasing returns to scale, there is a motivation to grow larger. As the larger a plant becomes, the closer it approaches maximum returns to scale and therefore the more important other considerations, such as location, become (Rawstron 1958).

Various case studies discuss the role of clusters to achieve rural economic development such as the work of Braden et al. (1998), Austin and Lozano (1999) to mention just two. Braden et al. (1998) analyze in detail particular cases of cluster occurrence in a resource-based industry and Austin and Lozano (1999) present an example for rural development based on a cluster strategy for Mexico's rural sector. However, these as well as other cases only present anecdotal evidence favoring industrial clustering. There is a need to develop formal mathematical models that help explain why resource-based industries may locate in certain areas and identify the causes behind such spatial arrangement for economic development and policy reasons.

1.4 Sawmill Lumber Location in the Wood Products Industry

Identification of a source of raw material that can supply sufficient inputs is essential to the development of a resource-based industry. In the wood products sector procurement functions and decisions are common to both small and large sawmills regardless of their total capacity. Procurement functions include: resource identification, purchasing, harvesting, resource allocation, receiving, inventorying, and primary processing (Harris 1988).

Identification of timber resources involves determination of quantity, quality, and species of the raw material. Once a source area has been identified, arrangements must be made to purchase the necessary quantity of timber. Methods of purchasing vary widely depending on the region's landownership pattern, competition among buyers, the sellers' plans, and tradition.

Timber can be purchased on a lump-sum basis, a per-unit basis, or in some combination. The harvesting component of wood procurement includes planning for and managing the harvest. It is usually done by contractors. During or following harvest, consideration must be given to allocating timber to its highest and best use (for example, timbers usable as sawlogs must be allocated to sawmills and pulpwood to pulpmills).

Transportation is an important cost element in all wood procurement systems. According to Harris (1988) procurement decisions are often subject to constraints imposed by the receiving mill, the terrain, or the regional infrastructure. Transportation decisions may be short-term in nature, such as scheduling decision for immediate delivery of logs to a mill, or they may be long-term, such as designing a road network, barge facilities or railroads.

Based on a survey among wood manufacturers, Barbe (1993) reports that truck transportation is the most commonly used method for transporting logs to sawmills in the U.S. South. Trucking is the preferred method because of ease of loading/unloading material and its capacity to transport all forms of wood products from tree-length pulp wood and saw logs to wood chips. The second most common mode of transportation is barging. Barge transportation is limited to high quality saw logs because of the elevated cost of this type of transportation. Hauling wood in the U.S. South by rail is restricted to short-wood pulpwood and chips for paper manufacturing. Table 6 presents unit costs, hauling distance, time it takes to transport logs from the woods to the sawmill, and percentages of wood hauled using different modes of transportation for a common softwood sawmill in the U.S. South as reported by Barbe (1993).

Log procurement systems differ because of complex decision-making between regions. Specific conditions are unique to each procurement region because of varying terrain and timber conditions, land ownership patterns, levels of competition, weather, government regulations, and

all other factors influencing business decisions. As business conditions change, procurement systems also change (Harris 1988).

Table 6. Average unit cost, hauling distance, time from woods to sawmill, and percentages of wood hauled for the major timber transportation modes used to haul timber to a sawmill in the U.S. South.

	Transportation		
	Truck	**Barge**	**Rail**
Cost	$0.08 per ton per mile	$9.34 (fixed fee)	$0.05 per ton per mile
Hauling distance	10-150 miles	50-500 miles	>100 miles
Time from woods to sawmill	1-3 hours	1-10 weeks	2-5 days
Percentage of wood hauled	87%	11%	2%

Source: Barbe (1993).

The study of location decision in a resource-based industry such as the softwood lumber industry then requires the explicit incorporation of a spatial dimension. A spatial analysis approach is necessary in order to incorporate the spatial components involved in a theoretical specification and also as consequence of the nature of data as will be discussed in subsequent sections. Empirical validation of new spatial concepts and models demands a statistical and econometric methodology that takes into account location and spatial interactions as it is stressed by Anselin (2000). Because of the importance of this dimension to the study of a resource-based industry, some concepts regarding the incorporation of space in formal models for analysis are reviewed in the next chapter. The development of spatial statistical techniques has been recent and made possible in part thanks to advances in Geographic Information Systems (GIS) technology. As part of the Literature Review a survey of key concepts regarding GIS that should establish a clear link between the former and spatial analysis is presented.

The Literature Review continues with a succinct summary of the principal theories that have examined the location of firms and their aggregation in space. I start with the early discussion of the Theory of Location by Ross, Renner and Isard and continue with the more recent New Economic Geography Theory of Krugman and Fujita, and Michael Porter's Theory

of Clusters. These theories constitute the Theoretical Framework for the research problem related to the study of location and spatial distribution of companies in the softwood sawmill sector. I also include a review of the theory behind decision-making and choice behavior. This will serve as a foundation to the study of how decision makers choose to locate resource-based companies in particular places.

In order to study the factors influencing the spatial distribution of enterprises in the softwood lumber sector a three-component research is presented in the Methodology section. The first component involves the survey of decision makers in the softwood lumber industry that will be asked about the importance of key location factors for their current sawmills as well as the perceived characteristics considered if they were to select a location for a new softwood lumber enterprise. This component includes the application of a Conjoint Analysis on respondents' preferences and the use of Factor Component Analysis to reduce the number of variables presented to respondents into a smaller set of explanatory variables to the location decision problem. The second component takes the information derived from the Factor Component analysis to develop a cross sectional analysis that will attempt to detect a cause-relationship effect between selected explanatory variables and the presence of sawmill enterprises. The third component studies the presence of centrifugal and centripetal forces suggested by the New Economic Geography and Clusters theory in industry localization and looks at two cross sections to see actual changes in the industry over time. While the first component looks at stated preferences in regard to the attributes that decision makers consider influence location of a sawmill the other two components of the research look at actual firm behavior in the way that the industry aggregates in space. The method proposed for the study of industrial location by itself constitutes a contribution to the emerging literature on spatial analysis.

CHAPTER 2. LITERATURE REVIEW

2.1 Spatial Data Analysis

Spatial data analysis extends and modifies standard statistical techniques so that data point locations and their arrangement are given greater importance in the analysis of results. It involves the analysis of a dataset that consists of geographically referenced attributes. According to Haining (1990) data sites are referenced so that the relative positions of sites are considered when building a model for the data and necessary for the assessment of different hypotheses concerning the spatial arrangement of data or some other non-spatial characteristics. Spatial analysis has seen a recent and rapid growth fostered by theoretical concerns as well as by the ability to apply models to newly developed georeferenced databases (Anselin and Getis 1992).

Spatial data analysis is comprised of three elements (Haining 2003). First, it includes cartographic modeling. Data is represented as a map and given a location value, often latitude and longitude coordinates. Second, spatial analysis incorporates forms of mathematical modeling where model outcomes are dependent on the form of spatial interaction between objects in the model, spatial relationships, or the geographical position of objects within the model. Third, spatial analysis includes the development and application of statistical techniques specifically for the analysis of georeferenced data.

According to Cressie (1993) Edmond Halley was the first to consider the spatial dimension of data. Halley (1686) superimposed, onto a map of land reforms, directions of trade winds and monsoons in three regions between and near the tropics, and assigned a physical cause to the winds. Student (1907) presented one of the first analyses where spatial information was formally incorporated into a model. Student studied the distribution of particles throughout a liquid and determined that one of the sources of error when counting yeast cells with a haemacytometer was that the distribution of cells over the area examined is not uniformly

distributed and it constitutes an error of random sampling. Later, Fisher realized the importance of spatial correlation when designing agricultural field experiments (Cressie 1993). Besag (1974) provides an example of a conditional probability model for spatially interacting random variables and suggests specific methods for statistical analysis of lattice systems.

2.1.1 Why Spatial Data Analysis?

The need to develop economic models that incorporate interaction with the local geography an environment has been discussed for long. Lindley Keasbey (1901) was one of the first to present a formal linkage between geography and economic systems. Keasby (1901) stressed that economic activities are first determined by the phenomena of nature, which is subsequently modified by human activities. Vaughn (1994) considers that in the new century the geography of resource economics is evolving into the proper study of behavioral economics and geography. This form of economic geography should be critical in aiding public policy particularly in regard to natural resource management.

Formal statistical analysis of spatially referenced data has received major attention in recent years and their application has gained wide acceptance as a methodology in mainstream sciences (Anselin 2000). Cressie (1993) presents a model for spatial data with a simple structure that can handle a large class of problems. Data may be continuous or discrete, there may be spatial aggregations or observations at points in space, spatial locations may be regular or irregular, and locations may be from a spatial continuum or a discrete set (Griffith 2000). Following Cressie (1993) let $\mathbf{s} \in R^d$ be a generic data location in d-dimensional Euclidean space and suppose that the potential datum $\mathbf{Z}(\mathbf{s})$ at spatial location $\mathbf{s}$ is a random quantity. Let $\mathbf{s}$ vary over index set R^d so as to generate the random process:

(1) $\{\boldsymbol{Z}(\mathbf{s}): \mathbf{s} \in D_r\}$

A realization of (1) is denoted $\{\mathbf{z(s)}: \mathbf{s} \in D_r\}$. D_r is a partitioned set of R^d with r=1,2,…, d such that for any r,s ($r \neq s$), $D_r \cap D_s = \varnothing$ and $D_r \cup D_s = R^d$. Cressie assumes that D_r is a random set, a measurable mapping from a probability space onto a measure space of subsets of R^d.

Cressie (1993) justifies the use of spatial methods for estimation, prediction and experimental design purposes. I will briefly refer to the first two issues. I will not touch on experimental design issues as this is not relevant to the purpose of this dissertation. For a complete discussion please see Cressie (1993).

- Estimation issues

Suppose $Z(1)$, … $Z(n)$ are independent and identically distributed observations drawn from a normal distribution with unknown mean μ and variance σ^2. The minimum-variance unbiased estimator of μ is

(2) $$\hat{\mu} \equiv \sum_{i=1}^{n} Z(i)/n,$$

and inference on μ is straightforward under the given assumptions: the estimator $\hat{\mu}$ is normally distributed with mean μ and variance σ^2/n. Hence, a two-tail 95% confidence interval for μ is

(3) $$\hat{\mu} - (1.96)\,\sigma/n^{½},\ \ \hat{\mu} + (1.96)\,\sigma/n^{½}$$

However, in the presence of positively correlated data, with correlation decreasing relative to the separation between observations, we have covariance

(4) $$\text{Cov}\,(Z(i), Z(j)) = \sigma^2 \cdot \rho^{\,|i-j|},\ \ i, j = 1, \ldots, n,\ 0 < \rho < 1\ .$$

This correlation function results from the first-order autoregressive process $Z(i) = \rho \cdot Z(i - 1) + \varepsilon(i)$, $i \geq 1$, where ρ is a parameter denoting linear correlation between observations, $\varepsilon(i)$ is an independent and identically distributed sequence of random errors with zero mean and variance $\sigma^2 (1 - \rho^2)$ and is independent of $Z(i - 1)$. Then, the variance for $\hat{\mu}$ is given by:

(5) $\mathrm{Var}(\hat{\mu}) = n^{-2} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{cov}(Z(i), Z(j)) \right\} =$

$= \sigma^2/n[\ 1+2\ \{\rho/(1-\rho)\}\{1-(1/n)\}\ -2\{\rho/(1-\rho)\}^2(1-\rho^{n-1}/n)]$.

Using a numerical example, Cressie (1993) illustrates how positive correlation in the data leads to a confidence interval that can be either too narrow or too wide resulting in biased estimates, thus, the need to consider the presence of spatial correlation to produce unbiased confidence intervals.

- Prediction issues

Cressie (1993) presents the problem of prediction in the presence of spatial correlation when an unknown observation $Z(n + 1)$ is to be predicted from data $\boldsymbol{Z} \equiv (Z(1), \ldots, Z(n))'$, where it is assumed $Z(1), \ldots, Z(n), Z(n + 1)$ are jointly Gaussian, identically distributed with unknown mean μ and known variance σ^2 and independent.

The predictor $p\,(\mathbf{Z};\, n + 1)$ that satisfies the unbiasedness condition $\mathrm{E}[p(\boldsymbol{Z};\, n+1)] = \mu$ and minimizes the mean-squared prediction error $\mathrm{E}[\boldsymbol{Z}(n + 1) - p(\boldsymbol{Z};\, n + 1)]^2$ is the sample mean

(6) $\hat{\mu} = p\,(\mathbf{Z};\, n + 1)$.

Under independence of the observations, the mean-squared prediction error is

(7) $ms_d = \sigma^2\ \{1 + (1/n)\}$.

When the independence condition is replaced by correlation between observations given by (4), then the new unbiased predictor that minimizes the mean squares error is

(8) $p_p\,(Z;\, n + 1) = \rho Z(n) + (1-\rho)\dfrac{\left\{ Z(1) = (1-\rho)\sum_{i=2}^{n-1} Z(i) + Z(n) \right\}}{(n-(n-2)\rho)}$.

When ρ is equal to zero, under independent observations condition, (8) reduces to (6). If the mean $\hat{\mu}$ is used instead of (8) the mean-squared prediction error is:

(9) $E(Z(n+1) - \hat{\mu})^2 =$

$$\sigma^2 \left\{1 + (1/n)[1 + 2\{\rho/(1-\rho)\}\{\rho^n - (1/n)\} - 2\{\rho/(1-\rho)\}^2 (1-\rho^{n-1})/n]\right\}.$$

Cressie (1993) indicates that the value of the mean squared prediction error between independent and dependent conditions is not large when spatial correlations decay geometrically with distance. According to Cressie, classical prediction intervals are often approximately valid but can be highly inefficient.

For a dependence model the mean squared predictor is

(10) $ms_p = \sigma^2 [1 - \rho^2 + \{(1-\rho)^2 \ (1+\rho)/(n - (n-2)\rho)\}]$.

If ρ is set to 0, then we obtain back the mean squared prediction error under independency (equation 7). For a large number of observations the squared ratio of the prediction-interval width from using $p_p(\mathbf{Z})$ to that from using $\hat{\mu}$ is approximately $1 - \rho^2$ which according to Cressie (1993) is a measure of the asymptotic efficiency of inference based on $\hat{\mu}$ versus the inference based on the optimal predictor $p_p(\mathbf{Z})$.

2.1.2 Spatial Dependence, Autocorrelation and Heterogeneity

Spatial econometrics deals with methodological concerns that follow from the explicit consideration of spatial effects such as spatial autocorrelation and spatial heterogeneity (Anselin 1999). Ignoring the effects of spatial effects in applied regression can result in severe errors in the interpretation of commonly used regression diagnostics and misspecification tests (Anselin and Griffith 1988). A concise review of such effects is presented in the following subsections.

2.1.2.1 Spatial Dependence or Spatial Autocorrelation

Spatial dependence or spatial autocorrelation refers to the lack of independence which is often present among observations in cross-sectional datasets (Anselin 1988). Spatial dependence refers to the relationship between spatially referenced data due to the nature of the variable(s)

under study and the size, shape, and configuration of spatial sampling units. The smaller the spatial unit the greater the probability that nearby units will be spatially dependent (Anselin and Getis 1992). Anselin (1988) considers that spatial dependency lies at the core of Tobler's first law of geography which states that "everything is related to everything else, but near things are more related than distanced things." Spatial dependence is determined by the notion of relative space or relative location which stresses the effect of distance.

Spatial dependence can be caused by various measurement problems common to applied work. Anselin (1988) mentions the arbitrary delineation of spatial units of observation such as census tracks or county boundaries, problems of spatial aggregation and more relevant to modeling the presence of spatial externalities, spillover effects as a consequence of measurement errors. The spatial organization and spatial structure of phenomena will tend to generate complex patterns of interaction and dependencies. Hence, models of spatial flows, spatial patterns, spatial structure and spatial processes should be able to capture elements of spatial dependence. In general terms, spatial dependency can be simply described as the existence of a functional relationship between what happens in one point in space and what occurs in all other points.

The existence of spatial dependence is often a cause of two major issues: the byproduct of measurement errors for observations in contiguous spatial units, and the existence of spatial interaction phenomena. Anselin (1988) presents a simple case to illustrate the presence of spatial dependency as a result of measurement errors. In empirical studies data is often collected at an aggregate scale (e.g. census information).There may be little correspondence between the spatial scope of the phenomenon under study and the delineation of the spatial units of observation and, thus, measurement errors are prone. Furthermore, measurement errors are likely to spill over across the boundaries of spatial units and errors for one unit of observation are likely to be linked to errors in a neighboring unit. Borrowing an example presented by Anselin (1988) let the true

spatial scales of the variables under study to be areas A, B and C (Figure 1). However, data collected is aggregated at levels 1 and 2, corresponding to observed variables Y_1 and Y_2. The result is that observed variable Y_1 is an aggregate of Y_A and part of Y_B. Observed variable Y_2 is an aggregate of Y_C and the remainder of Y_B.

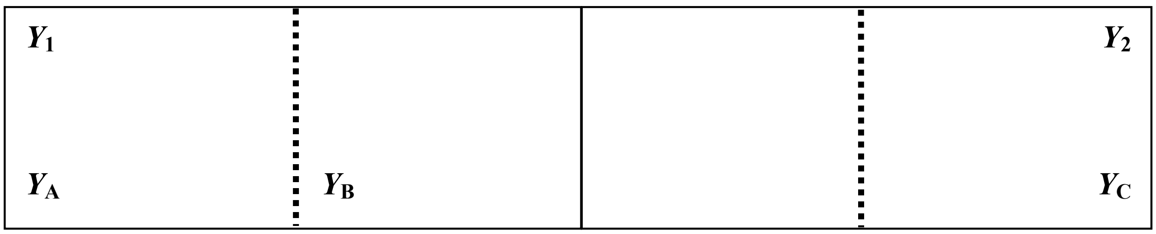


Figure 1. Illustration of spatial dependence as a result of measurement errors due to aggregation.
Source: Anselin (1988).

Formally:

$Y_1 = Y_A + \lambda Y_B$

$Y_2 = Y_C + (1 - \lambda)\, Y_B.$

This aggregation is likely to suffer from errors in the assessment of the weighting parameter λ, which is present in Y_1 as well as in Y_2. The result is a pattern of measurement errors that exhibits spatial dependence.

The other factor that causes spatial dependence is the existence of spatial interaction phenomena (Anselin 1988). Spatial interaction theories, diffusion processes, and spatial hierarchies result in formal models to structure the dependence between phenomena at different locations in space. Thus, what is observed at one point is a function of points elsewhere in the system. It can be formally expressed as:

$y_i = f(y_1, y_2, \ldots, y_N)$

where observations of variable y at point $i \in \mathbf{S}$. $\mathbf{S}$ is the dataset containing all spatial units of observation and is a function to the magnitudes for the variable in all other spatial units in the

system. This expression, in an empirical application, results in an unidentifiable system with many more parameters (as many as $N^2 - N$) than observations (N). By imposing a functional structure on the relationships for the spatial process of a limited number of characteristics of the spatial dependence process may be estimated and tested empirically (Anselin 1988). A more general representation of a spatial structure with dependent error terms is given by Anselin and Griffith (1988) and is expressed as:

$y = f(y, \mathbf{X}, \beta, \varepsilon)$, or

$\varepsilon = g(\varepsilon, \lambda, \xi)$,

where y is the dependent variable, $\mathbf{X}$ is a set of explanatory variables, β and λ are vectors of parameters and ε and ξ are error terms. The functional forms represented by f and g indicate how events of y or ε at one point in space are related to the points in the entire system. Frequently, the interaction between the value at one location and values at the remaining locations of the system takes the form of a weighted sum, such as $\mathbf{C}y$ or $\mathbf{C}\varepsilon$. The elements of the weight matrix $\mathbf{C}$ represent the potential spatial interaction between the data points and are often presented in binary contiguity form (Anselin and Griffith 1988).

2.1.2.2 Spatial Heterogeneity

According to Anselin (1988) spatial heterogeneity pertains to the lack of stability over space of the relationships under study. Precisely, this refers to the structural instability in the form of non-constant error variances (heteroskedasticity) or model coefficients (variable coefficients, spatial regimes) that vary with location and are not homogeneous throughout the dataset. In addition to a lack of structural stability spatial heterogeneity emerges due to the heterogeneity found within spatial units of observations. This is the case of census tracts where population or income levels are not homogeneously distributed across space, or where different

tracts have different areas and shapes. The reflection of these issues in measurement errors such as missing variables or functional misspecification can result in heteroskedasticity.

These aspects of spatial heterogeneity (structural instability and changing functional forms) can be illustrated in a regression context where cross-section data is combined with time series data for a general expression:

$y_{it} = f_{it} (x_{it}, \beta_{it}, \varepsilon_{it})$,

where index i refers to a spatial unit of observation and t to the time period. The function f_{it} is a time-space specific relationship which explains the value of the vector of dependent variables y_{it} in terms of a vector of independent variables x_{it}, a vector of parameters β_{it}, and an error term ε_{it}. This formulation is not operational as there are more parameters than observations. In order to carry out estimation and inference there is a need to apply insights from spatial econometric theory to impose a spatial structure and spatial interaction as the basis for the various constraints and reparameterizations. For models with varying coefficients, as is represented by β_i above, variation should be determined systematically as a function of a small number of additional variables, or stochastically, in terms of a priori distribution. In the case of structural instability of the function form, such as in the form of f_i, the number of different regimes that can be efficiently estimated is limited by degree of freedom considerations (Anselin and Griffith 1988).

Anselin (1988) stresses that it is important to consider spatial heterogeneity explicitly for three reasons:

- The structure behind the instability is spatial (geographic). The location of the observations is crucial in determining the form of the instability. For example, groupwise heteroskedasticity could be modeled as different error variances for different geographic subsets of the data. Spatial groupwise heteroskedasticity would follow in the form of spatially clustered error variances for observation i, $\mathrm{Var}[\varepsilon_i]=\sigma^2_r$ when $i \in D_r$.

- Because the structure is spatial, heterogeneity often occurs jointly with spatial autocorrelation, and standard econometric techniques are no longer appropriate.
- In a single cross-section, spatial autocorrelation and spatial heterogeneity may be observationally equivalent. For example, a spatial cluster (i.e., observed in locations that are in close proximity) of extreme residuals may be interpreted as due to spatial heterogeneity (e.g., groupwise heteroskedasticity) or as due to spatial autocorrelation (e.g., a spatial stochastic process yielding clustered values). This requires that both aspects of the problem be structured very carefully to obtain identifiability of the model parameters, and that one aspect can never be considered in isolation from the other.

Anselin (1988) suggests that problems of spatial heterogeneity are likely to occur in econometric models estimated on cross-sectional data sets of dissimilar spatial units.

2.1.3 Spatial Regression Models

According to Anselin (2001) a spatial component is incorporated into an econometric model as a proxy for missing variables or other forms of specification imperfections. The scale mismatch and the inherent need to integrate data from various sources will tend to result in spatially dependent as well as spatially heterogeneous observations.

One of the challenges faced in spatial modeling is the formal incorporation of a spatial dependence structure in the model. A spatial dependency structure is usually presented as an analog to time series models that incorporate a serial correlation structure and lagged variables. Nevertheless, standard econometric approaches applied to time series do not carry over in a straightforward fashion to spatial dependence in cross-sectional samples (Anselin 1988). This is a direct result of the multidirectional dimension of dependence in space which differs from the unidirectional structure of time series. Furthermore, in spatial processes, nonlinear maximum likelihood estimation is necessary whether or not the error term is dependent. Contrary, in the

absence of serial correlation in the error term, ordinary least squares can be used and would result in consistent parameter estimates in a time series context (Anselin and Griffith, 1988).

In the traditional linear regression model spatial dependence can be incorporated in two distinct ways: as an additional right-hand side component in the form of a spatially lagged dependent variable (**W**y) or in the error structure (E[ε_i, ε_j] $\neq 0$). According to Anselin (1988) the spatial lagged model is appropriate when focus is on the assessment of the presence and strength of spatial interaction. On the other hand, spatial dependence in the error term (also referred as nuisance dependence) is suitable when correcting for the potential biasing influence of spatial autocorrelation due to the use of spatial data, regardless of whether the real model is spatial or not.

As presented by Anselin (1988) spatial autocorrelation can be expressed by the moment condition:

(11) $\text{Cov}[y_i, y_j] = \text{E}[y_i, y_j] - \text{E}[y_i] \cdot \text{E}[y_j] \neq 0$, for $i \neq j$,

where i, j correspond to individual observations. The covariance matrix in (11) has an important interpretation in terms of spatial structure, spatial interaction and spatial arrangement of the observations. **W** is a n x n spatial weights matrix. By convention, the values of the diagonal of the spatial weight matrix w_{ii} are set to 0 so that a data is not a neighbor of itself. Further, the elements of the spatial weights matrix are typically row standardized such that for each y_i the sum of all w_{ij} values is equal to 1. Then, the spatial lag can be interpreted as a weighted average of the neighbors as will be presented next, or what Anselin (1988) calls a spatial smoother.

The notion of spatial dependence involves the need to identify which other units in the system have an effect on the value of a particular unit under consideration. Anselin (1988) considers this issue to be expressed in the topological notions of neighborhood and nearest neighborhoods. Let a spatial system **S** contain Ni units (i=1,2,…N) and a variable y observed at

each of the spatial units. In lattice processes a set of neighbors for the spatial unit i is defined as the compilation of those units j for which y_j contained in the functional form of the conditional probability of y_i, conditional upon x at all other locations. Anselin suggests a formal definition that would yield the set of neighbors for i as J, for which:

$P[y_i \mid y] = P[y_i \mid y_j]$.

Where y_j is the vector of observations for $y_j \;\forall\; j \in J$, and y is the vector containing all observed y values in the system. Alternatively, Anselin (1988) suggests that the set of neighbors j for i can be taken as those locations for which the conditional marginal probability for x_i is not equal to the unconditional marginal probability. Formally,

(12) $\{j \mid P[x_i] \neq P[x_i \mid x_j]\}$.

The above definitions do not include information about the relative location of two spatial units but only refers to how they influence another via conditional probabilities. Anselin (1988) adds a spatial aspect to these definitions in the following expression:

(13) $\{j \mid P[x_i] \neq P[x_i \mid x_j]$ and $d_{ij} < \epsilon_i$,

where d_{ij} is a measure of the distance between i and j in space, and ϵ is a critical cut-off point for each spatial unit y_i and, similarly for all spatial units. The distance metric underlying d_{ij} can refer to a Euclidean, Manhattan Block or general Minskowski distance. This notion of neighbor introduces an additional structure in the spatial dataset. It now combines a statistical dependence dimension (relating magnitudes at different observation units) with a notion of space (distance and relative location). Nevertheless, this definition does not exclude spatial units j that are located beyond the cut-off point d_{ij} to influence the conditional probability of y_i even though these are not considered to be neighbors. Excluded points can have an influence on y_i via other spatial units which correspond to the concept of higher order neighbors (Anselin 1988).

- <u>Spatial Lag Model and the Simultaneous Approach</u>

According to Whittle (1954) spatial lag processes, compared to time series, are influenced by observations in all directions and not only past values. This results in a deviation from the standard regression model

(14) $y = \mathbf{X}\beta + \varepsilon$

where y is a $n \times 1$ vector of dependent random variables in the R^d space, $\mathbf{X}$ is a $n \times k$ matrix of explanatory variables, β is a $k \times 1$ vector of regression parameters and ε is a $n \times 1$ vector of random errors.

A spatial lag model involves the use of a weighted average of random variables at geographically closed locations. In this concept the spatial weight matrix $\mathbf{W}$ defines the observations that affect the dependent variable y as the values in the weight matrix that are different from zero. A spatial lag for y at point i can then be expressed as

(15) $\mathbf{W}y = \sum_{i=1}^{n} w_{ij} . y_j \, , j = 1, \ldots \mathrm{n}$

where y is a n by 1 vector of observations on the random variable. A spatial lag model then is estimated by adding the weighted average of the value of y in the neighboring locations to the right-hand side of the regression model (14). It yields

(16) $y = \rho\mathbf{W}y + \mathbf{X}\beta + \varepsilon$

where ρ is the spatial autoregressive coefficient and the other components are as defined previously. The most common approach to the spatial lag problem is the simultaneous model of spatial dependence (Anselin 2001, Goodchild and Haining 2004). In the simultaneous spatial autoregressive model the process can be represented as

(17) $(y - \mu_i) = \rho\mathbf{W}(y - \mu_i) + \varepsilon,$ *or* $(y - \mu_i) = (\mathbf{I} - \rho\mathbf{W})^{-1}\, \varepsilon.$

In this approach the full array of the variable of interest is determined by the model. The variance-covariance matrix for y is determined by the constraints imposed by the weight structure of **W** and the specific (exogenously determined) form of the spatial process. The result is that the variance-covariance matrix is a function of two parameters, the variance σ^2 and the spatial coefficient ρ. For the simultaneous autoregressive process, since $E[y- \mu_i] = 0$, the variance covariance matrix is defined by (Anselin 1999, Goodchild and Haining 2004) as

(18) $\text{Cov}[(y - \mu_i), (y - \mu_i)] = E[(y - \mu_i)\,(y - \mu_i)'] = \sigma^2\,[(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}$

which is a full matrix. This structure implies that a shock at any location affect the entire set of locations through what Anselin (1988) calls a spatial multiplier effect.

Anselin (1999, 2001) stresses that despite the error terms are considered to be independent and identically distributed, the spatial lag term is also correlated with them as presented in the reduced form of (16),

(19) $y = (\mathbf{I} - \rho\mathbf{W})^{-1}\,\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon.$

- Spatial Error Model

The error term in a regression model accounts for disturbance that arises because of omitted variables, incorrect functional form, and errors of measurement (Greene 2003). If the error at each location depends on the errors at other locations, then it is said that the errors are spatially autocorrellated. In this case, the assumptions of homoskedastic and uncorrelated errors are not satisfied (Anselin and Hudak 1992). There are many forms of spatial dependence in the error term, but usually, a spatial autoregressive process for the error term is estimated. This model is the standard regression specification with a spatial autoregressive error term:

(20) $y = \mathbf{X}\beta + \varepsilon$

$\varepsilon = \lambda\mathbf{W}\varepsilon + \gamma$

where y is an N by 1 vector of observations on the dependent variable, **X** is an N × k matrix

of observations on the explanatory variables, β is a k × 1 vector of regression coefficients, ε is an N × 1 vector of error terms, $\mathbf{W}\varepsilon$ is a spatial lag for the errors, λ is the autoregressive coefficient and γ is the error vector with a mean of zero and constant covariance $\sigma^2\mathbf{I}$.

Florax and Folmer (1992) mention the importance of the specification of the **W** matrix. Misspecification of the weight matrix has an impact on hypothesis testing with respect to spatial dependence among residuals as well as a drawback in terms of bias regarding the estimated coefficients in models with spatially lagged variables. Also the power of autocorrelation statistics such as Moran's I and Lagrange multiplier statistics are critically dependent on the type of weight matrix and whether this has been standardized. These tests will be introduced in the Methods sections.

2.1.4 Spatial Model Estimation

Spatial dependence among disturbances of spatial models is a serious problem in empirical research (Florax and Folmer 1992). Specifically, commonly used Ordinary Least Squares (OLS) estimator is inefficient, the estimator of the residual variance is biased, the values of the estimated R^2 are inflated and inference procedures are invalid (Cliff and Ord 1981). Besides, the presence of spatially correlated residuals affects the properties of tests for model selection and heteroskedasticity (Anselin and Griffith 1988).

In the mixed spatial autoregressive model OLS is inconsistent due to the presence of the spatial lag $\mathbf{W}y$ on the right hand side of the regression (Anselin 1988). This situation is similar to the presence of endogenous variables in a system of simultaneous equations. It is for this reason that this model is known as simultaneous spatial autoregressive model (Anselin and Hudak 1992). The most commonly suggested approach for estimation is the use of maximum likelihood.

For the spatial autoregressive error model, a special case of regression with a non-spherical error term (errors that meet the characteristics of homoskedasticity and non-

autocorrelation are often called spherical disturbances, Greene 2003), OLS estimates are still unbiased but are no longer efficient. Therefore, inference based on the biased OLS estimates for variance and model fit may be misleading. This is because the error variance for this model is

(21) $E[\varepsilon\varepsilon'] = \sigma^2[(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}$.

According to Anselin (1988) there is no consistent two-step or iterative estimator for this model, the only alternative is maximum likelihood estimation. Table 7 presents a list of the most common spatial regression approaches, along with their appropriateness in regard to the research problem and suggested estimation methods.

Table 7. Summary of most common spatial regression approaches

Regression Model	**Appropriateness**	**Estimation Method**
Classical regression $y = \mathbf{X}\beta + \varepsilon$ $\varepsilon \sim N(0, \sigma^2 I)$	Independent and identically distributed. observations	Ordinary Least Squares
Spatial autoregressive lag model $y = \rho\mathbf{W}_1 y + \mathbf{X}\beta + \varepsilon$ $\varepsilon \sim N(0, \sigma^2 I)$	Focus of interest is the assessment of the existence and strength of spatial interaction.	Maximum Likelihood Estimation, Spatial Two Stage Least Squares
Spatial autoregressive error model $y = \mathbf{X}\beta + \varepsilon$ $\varepsilon = \lambda\mathbf{W}_1\varepsilon + \gamma$ $\gamma \sim N(0\ \sigma^2 I)$,	Appropriate when the concern is with correcting for the potentially biasing influence of the spatial autocorrelation due to the use of spatial data	Maximum Likelihood Estimation, Method of Moments
Spatial moving average model $y = \mathbf{X}\beta + \varepsilon$ $\varepsilon = \lambda\mathbf{W}_1\gamma + \gamma$ $\gamma \sim N(0, \sigma^2 I)$	Focus of interest is the assessment of the existence and strength of spatial interaction.	Maximum Likelihood Estimation, Method of Moments
Mixed regressive-spatial autoregressive model $y = \rho\,\mathbf{W}_1 y + \mathbf{X}\beta + \varepsilon$ $\varepsilon = \lambda\mathbf{W}_2\varepsilon + \gamma$ $\gamma \sim N(0, \sigma^2 I)$	Appropriate in the presence of strong spatial interaction and potential bias caused by spatial data.	Maximum Likelihood Estimation, Spatial Two Stage Least Squares
Mixed regressive-spatial regressive model with autoregressive disturbances $y = \zeta\mathbf{W}_1 y + \mathbf{X}\beta + \mathbf{W}_2\mathbf{X}^*\rho + \varepsilon$ $\varepsilon = \lambda\mathbf{W}_3\varepsilon + \gamma$ $\gamma \sim N(0,\Omega)$, and the diagonal elements of the error variance matrix Ω as $\Omega_{ii} = h_i(z\alpha)\ \ h_i > 0$	Appropriate in the presence of strong spatial interaction and potential bias caused by spatial data	Maximum Likelihood Estimation, Spatial Two Stage Least Squares

Where y is a N by 1 vector of observations on the dependent variable, $\mathbf{X}$ represents a N by K matrix of observations on the explanatory variables, ε represents a N by 1 vector of error terms, ρ, ζ and λ are spatial autoregressive coefficients.
Source: Anselin (1988), Anselin and Hudak (1992), Florax and Folmer (1992).

A larger variety in the taxonomy of spatial models can be obtained by introducing different weight matrices for example we may have a model $y = \zeta \mathbf{W}_1 y + \mathbf{X}\beta + \mathbf{W}_1\mathbf{X}\rho + \lambda\mathbf{W}_2\varepsilon + \gamma$ where $\mathbf{W}_1 \neq \mathbf{W}_2$. It is important to note that the main difference between moving average and autoregressive processes is that in the case of the latter spatial effects extend to all locations in the system, while spatial effects are limited to first and second order in the moving average process (Florex and deGraaff 2004).

2.1.5 Application of Spatial Regression Models

The literature on the incorporation of a spatial dimension to regression models applied to natural resource problems is relatively new. Smith (2002) presents a study of the spatial behavior of renewable resource harvesters for the fisheries industry. One of Smith's approaches to analyze this problem was to assume a random utility model for the decision to harvest. Smith aggregated observations for each harvester across space and used a Nested Logit Regression to model this problem. The Nested Logit was used because it allows for individual heterogeneity and does not suffer from the independence of irrelevant alternatives restriction of other discrete models (i.e. basic multinomial or logit). Furthermore, Nested Logit Model is simple to use for policy simulation because it does not require calculation of an individual agent's entire optimal path nor integration over individual heterogeneity that is manifested in random parameters (Smith 2002).

In another example with a direct link to a natural resource Cohen and Paul (2005) studied agglomeration economies and their effect on industry location decision for the food processing sector in the U.S. Cohen and Paul (2005) used a spatial autoregressive model to analyze agglomeration in the food manufacturing industry. Cohen and Paul follow Anselin (1988) and introduce a spatial dimension in their analysis through the inclusion of a weight matrix that gives all neighboring states equal weight and all other states zero weight. Cohen and Paul (2005) found that significant average cost-savings occur by state-level food manufacturing industries from

spatial agglomeration within a state. Contrary, marginal costs are higher in areas of high consumer demand, due to congestion or quality impacts, but lower in rural areas.

Previous research has incorporated a spatial component in a model for productivity. Ciccone and Hall (1996) present a model that attempts to include such spatial dimension through the development of an index for economic activity per unit of geographic area. Specifically, Ciccone and Hall studied the relation between productivity and density of economic activity in the U.S. According to Ciccone and Hall (1996), spatial density of economic activity is the main source of aggregate increasing returns. Higher density of economic activity then results in increasing returns, external economies and a higher degree of production specialization.

But application of spatial econometric models to the forest sector has been rather limited. Blackman et al. (2003) present what they consider to be the first spatial regression analysis of land cover changes in a managed forest ecosystem. Blackman et al. (2003) used high-resolution land cover data derived from aerial photographs, along with data on the institutional, geophysical, socioeconomic, and agronomic characteristics of the study area. They conclude that plots in close proximity to urban centers are less likely to be cleared, all other things equal. They also consider that membership in marketing cooperatives, farm size, and certain soil types are associated with forest cover, while common property, proximity to small town centers, and the prevalence of indigenous peoples are associated with forest clearing. Blackman et al. (2003) corrected for spatial autocorrelation using a Bayesian heteroskedastic spatial autoregressive procedure for a logit model (LeSage 2000).

Munroe et al. (2004) studied land cover changes in western Honduras using satellite images. They used two econometric specifications, multinomial logit and binary logit, to model the likelihood of changes in land uses. Munroe et al. (2004) follow Nelson and Hellerstein (1997) to incorporate a spatially weighted average of slop at neighboring locations. Munroe et al.

(2004) justify the use of a spatial regression model because they consider these effects to the underlying behavior model. Because of the use of remote sensing data are likely to be highly correlated. The use of GIS coverages is subject to measurement errors and any imprecision will likely yield spatially correlated errors. In addition to biophysical processes, such as vegetation or climate, many human activities exhibit spatial correlation. For example, land cover changes tend to occur in areas proximate to clearing from a previous period (Pinkse and Slade 1998).

Blackman et al. (2006) make use of a Probit model to analyze the likelihood that a land plot has been deforested when studying shade-grown coffee areas in El Salvador, Central America (Blackman et al. 2006). They used regression analysis to identify the key drivers of tree cover loss in El Salvador's coffee areas between 1990 and 2000. They used a series of probit regressions in which the dependent variable was a measure of tree cover change between 1990 and 2000 on each plot, and the independent variables were proxies for, or measures of, the geophysical, institutional, and agronomic, and socioeconomic characteristics of each plot.

Aguilar and Vlosky (2006) present an analysis of forest product manufacturer clusters in Louisiana. This analysis explores the spatial distribution of primary and secondary forest products manufacturers in Louisiana in order to identify spatial clusters and model industry frequencies as a function of socio-economic variables. They develop a spatial correlated error regression that models the number of manufacturers per zip code block as a function of several explanatory variables. Results suggest that primary forest products companies show a higher spatial dependency compared to secondary forest products manufacturers as well as evidence of clustering of secondary forest products manufacturers. Regression coefficients show that total population is the variable most significantly correlated to clustering of secondary forest products manufacturers.

2.2 Geographic Information Systems (GIS)

A Geographic Information System (GIS) can be defined as a computer-based information platform which attempts to capture, store, manipulate, analyze and display spatially referenced and associated tabular attribute data (Fischer and Nijkamp 1992). GIS is a data processing system that can be used as technical support to facilitate spatial analysis and decision making. GIS can be distinguished from information systems used for business data processing by its ability to store, handle and analyze georeferenced data.

GIS makes use of spatial datasets that describe locational positions of objects in spatial or space-time systems and maintain non-spatial information (attributes) of the objects recorded. The system is comprised of (1) a database of spatially referenced data, (2) appropriate software components required for data manipulation and analysis, and (3) necessary hardware components for displays and electronic storage.

Fischer and Nijkamp (1992) consider that the function of an information system is to assist a user in solving complex research, planning and management problems and by improving the user's ability to evaluate policy issues, compare alternatives, and ultimately facilitate the decision-making process. Figure 2 summarizes the components and capabilities of a GIS.

Early forms of automated cartography appeared in the late 1950's with advances in computing technology (Goodchild and Haining 2004). Later, with the development of scanners, plotters and software, new possibilities for spatial display and early spatial analysis emerged in the 1960s. In 1963 Matheron published his method for point interpolation. This technique is applied to the case of spatial processes defined on continuous geographic space and is commonly known as Kriging.

The first commercially viable GIS appeared in the early 1980s, as the advent of the personal computer made it possible to acquire sufficient dedicated computing power and as

relational database management software obviated the need to construct elaborate data-handling functions (Goodchild and Haining 2004). Since the National Science Foundation created the National Center for Geographic Information and Analysis in 1998 it triggered the rapid growth and development in the use and applications of GIS (Getis 2000).

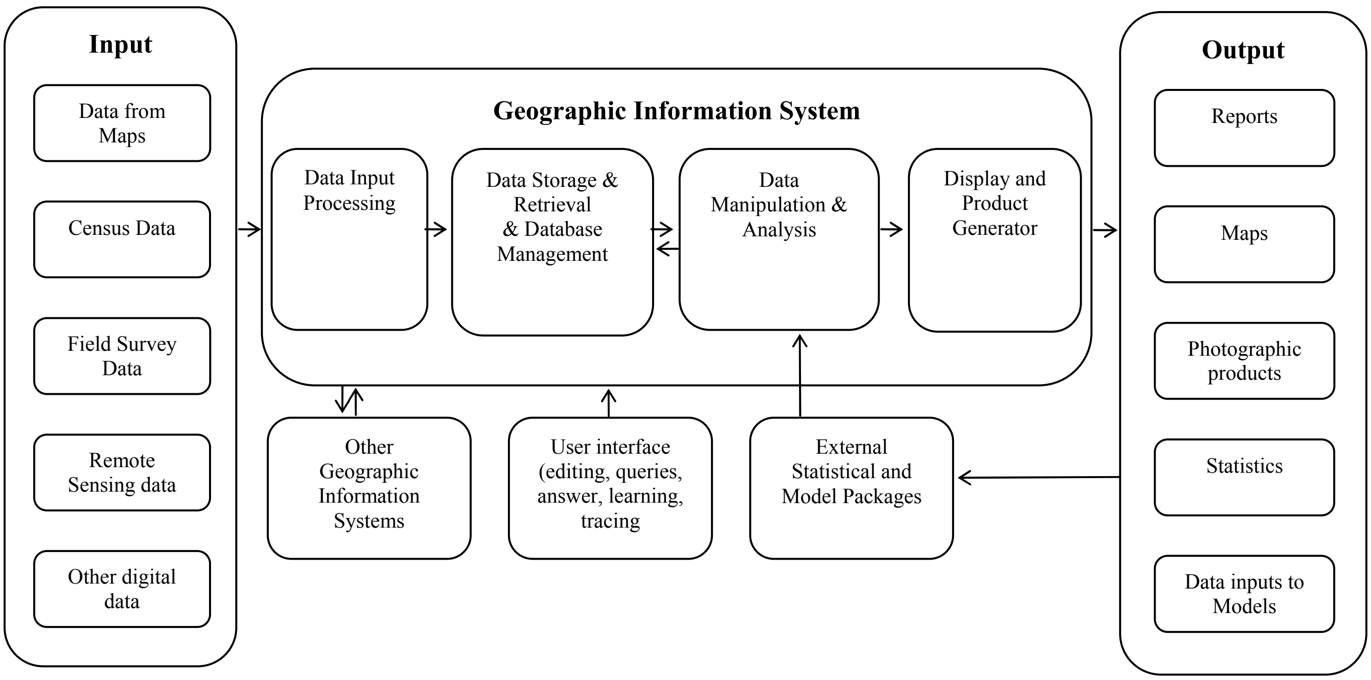


Figure 2. Components and capabilities of a Geographic Information System (GIS).
Source: Fischer and Nijkamp (1992).

Historically, GIS has focused on measuring the location of points on the earth's surface in absolute terms (latitude and longitude coordinates). While this has been appropriate for mapping applications it is not adequate for many uses in social sciences. Goodchild and Haining (2004) argue that in social sciences it is the relative position that is important, such as the distance between observation points.

GIS and spatial data analysis meet at the spatial data matrix used in a regression model (Goodchild and Haining 2004). The data matrix should be able to capture the structure of the spatial process taking place. At a conceptual level, this matrix consists of rows and columns

where rows refer to cases and columns refer to the different attributes measured at each of the cases. The last columns of the matrix provide spatial referencing information for each observation (i.e. latitude and longitude coordinates). For example, following Haining (2003), let $Z_1, Z_2, \ldots Z_k$ represent the k variables or attributes and **S** to location information. Then a spatial data matrix can be presented as in the following matrix:

$$\begin{pmatrix} z_1(1) & z_2(1) & \ldots & z_k(1) & \tilde{s}(1) \\ z_1(2) & z_2(2) & \ldots & z_k(2) & \tilde{s}(2) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ z_1(n) & z_2(n) & \ldots & zk(n) & \tilde{s}(n) \end{pmatrix}$$

where lower cases for Z denote an actual data value and the numbers 1,2,…,n references a particular observation. Attached to every case (i) is a location $\tilde{s}(i)$ which indicates this is a vector and may contain more than one number for the purpose of identifying the spatial location for case i such as $s(i)=(s_1(i),s_2(i))$.

According to Fischer and Nijkamp (1992) spatial analysis can take advantage of the rich amount of data from new sources available in geographic information systems. GIS can generate information that can be included in regression models that before the advent of georeferenced data was complex to obtain. For example, some spatial operations based on spatially referenced data include (Fischer and Nijkamp 1992):

- Geometric calculation operators such as distance, length, perimeter, area, closest intersection and union.
- Topological operators such as neighborhood, left and right polygons of a polyline, start and end notes of polylines.
- Spatial comparison operators such as intersects, inside, larger than, outside, neighbor of.
- Multilayer spatial overlay involving the integration of nodal, linear and polygon layers.

CHAPTER 3. THEORETICAL FRAMEWORK

This chapter presents the theoretical framework on which this dissertation expands on. This section starts with the presentation of the general Location Theory developed since the early 1900s that is followed by emergence of the New Economic Geography. The theory of Clusters then introduces several factors that facilitate and motivate industrial agglomeration. Finally, a review of Choice theory and random utility is presented to illustrate the decision making process behind the choice of location.

3.1 Location Theory

Renner (1931) formulated a general Principle of Industrial Location which states that: "*Any industry tends to locate at a point which provides optimum access to its component elements. If all these component elements are juxtaposed, the location of the industry is predetermined. If, however, they occur widely separated, the industry is so located as to be most accessible to that element which would be the most expensive or difficult to transport and which, therefore, becomes the locative factor for the industry in question*."

Alfred Weber's "Theory of the Location of Industries" attempts to present an isolating analysis that would help identifying causal relationships and develops laws of industrial location within a territory which represents a politically and nationally uniform organization (Weber 1929). Weber presented his work as a complement to von Thünen's Theory of Location of Agricultural Production. In Weber's analysis, the forces which operate as economic causes of industrial location are known as locational factors, advantages which are gained when an economic activity takes place at a particular location or at several such locations because the entire productive and distributive process is completed at a lower cost than elsewhere.

Weber (1929) recognizes factors that influence location as being general or special in nature. General factors are considered in the case of every industry when deciding a place to

locate and include costs of transportation, labor and rent. Special factors refer to conditions of concern to particular industries only such as perishability of raw materials, the influence of degree of humidity on the manufacturing process or dependence of the manufacturing process upon water quality.

Weber further classifies general or special conditions into regional and agglomeration/deglomeration factors. Regional factors also referred to as direct or primary, draw industries to definite regions because of cost advantages (Renner 1931). Agglomerating and deglomerating factors are defined by Weber (1929) as specific economic considerations that particular industries reflect on within a specified subregion. Generally speaking, an indirect factor is an advantage which follows from the fact that not less than a certain quantum of production is agglomerated at one place (agglomerating factor), or from the fact that not more than a certain quantum of production is agglomerated at one place (deglomerating factor). Agglomerating factors derived from advantages from large-scale production are related to the nature of the particular industry, while the deglomerating factors are all linked to the inevitable increases in land rent which accompany industry agglomeration. Weber also mentions natural, technical, social and cultural conditions that can draw industries closer or thither.

Alfred Weber (1929) enumerates seven cost elements affecting the manufacturing of goods. These are cost of grounds, cost of buildings, machines and other fixed capital costs, cost of securing materials, power and fuel, cost of labor, cost of transportation, interest rates, and rate of depreciation of fixed capital. Of the above mentioned elements Weber argues that only four of them vary according to the location of the place of production and, thus, represent general factors of location. These factors are (1) costs of buildings, machines, and other fixed capital costs, (2) costs of securing materials, power, and fuel, (3) costs of labor, and (4) costs of transportation.

In general Weber's type of locational analysis focused on the cost minimizing conditions required in order to produce and ship goods from the manufacturing site to their consumption market, while including transportation costs incurred in the delivery of required inputs (McCann and Sheppard 2003). In order to analyze the main factors influencing the geographical distribution of manufacturing industries Weber (1929) imposes several assumptions. Weber assumes that geographical basis of raw materials and places of consumption are given, labor is not mobile, wages of each brand of industry are fixed and labor availability is unlimited. Under this framework of analysis the most important factors determining the spatial orientation of industries within a region are transportation and labor costs. The level of deviation from the minimum transportation cost figure is given by the total amount of labor required per ton per product.

Transportation costs define industrial orientation depending on the type of transportation used, nature of the location and its kind of roads, nature of goods themselves and qualities that determine ease of transportation. Weber (1929) introduces a "material index" of production, which refers to the proportion of the weight of used localized input materials to the weight of the manufactured product. This index indicates how many weight units of localized material have to be moved in addition to the weight of the product. The total weight to be moved in a locational figure per unit of product helps determine the locational weight of the particular industry. Then, industries having a high locational weight (high material index) are attracted toward the location of materials and industries having low locational weight are attracted towards consumer markets. In other words, production of goods using high weight-losing raw materials pulls manufacturing to their input deposits, while manufacturing of goods using low-weight using inputs will occur near the place of consumption. According to Weber (1929) the only factor that affects choice of location, as far as transportation is concerned, is the material index of the industry.

Labor costs levels are assumed to be fixed and the supply of labor is unlimited in Weber's location theory analysis. Labor costs become a factor to industry orientation only when they differ from place to place. Labor costs, in the form of wages, vary because of differing levels of efficiency and differences in the use of technical equipment with which the laboring force is set to work at particular locations.

Combining transportation and labor costs, a manufacturing location can be moved from the point of minimum transportation costs to a more favorable labor location only if the savings in the cost of labor, which this new place makes possible, are larger than the additional costs of transportation which it involves (Weber 1929).

Industrial agglomeration results from the social nature of the production process. An agglomerative factor is a cost advantage of production or marketing which results from concentrating production to some considerable extent at one place, while a deglomerative factor refers to the cheapening of production thanks to the decentralization. According to Weber (1929) agglomerative factors occur at the plant level (concentration of industry through the simple enlargement of plant until a minimum efficient size is reached) and by the association of several plants. Various plants may decide to cluster together to maximize the use of product specific technical equipment, the development of a large pool of labor, take advantage of marketing factors (i.e. buy and sell on a large scale), and reduce overhead costs. Deglomerative factors, on the other hand, have a counteractive effect and are most commonly derived from the rise of land value that often results in the increase of overhead costs.

But it was probably Ross (1896) the first to distinguish between the process of concentration and migration in the location of industries. Ross considers that an industry initially scattered geographically concentrates in one place because of economies that result from dwelling of many enterprises of one kind in the same neighborhood. For example, a large

number of work-people living in a single community can develop skills much faster than if scattered in small individual groups. Working methods can be leveled up to the best known, making it easier for a new process or skill to become a standard.

Location privileges attract industries (Ross 1896). Among the factors that magnetize industries Ross mentions easy incorporation, light taxes, severe penalties for offenses against property, generous grants of authority to private watchmen, flexible legislatures and complaisant courts. Renner (1931) mentions factors such as adverse or favorable laws, taxation policies, climatic conditions, industrial supplies, facilities for waste disposal, local pride and encouragement, policies of labor unions, and cost of living.

The concentration of an industry results from the advantages found in one locality that outstrips others. As a center leaves its rivals behind and industry agglomerates in one location, economies of concentration emerge favoring its growth. But growth is not infinite. Industry concentration growth is restricted because the special advantages that caused industry concentration in the first place tend to disappear as input and transportation costs rise (Ross 1896). The multiplication of enterprises demands more inputs, raising their costs, ultimately making them no longer cheaper than elsewhere. Also, as industry concentrates, the radius of the territory from which its materials and the subsistence of its dependent population are drawn, and of the territory over which the finished product is distributed increases, the average cost of transportation per unit of industry grows, until its growth neutralizes the economies of further concentration. Renner (1931) cites urban congestion, social problems, higher rents, burdensome taxes, mounting insurance rates and inability to sustain working populations through depressions as forces contributing to decentralization.

Helburn (1943) criticizes the use of the word location in economic geography and rather suggest the distinction between three different forms of “location”: industry orientation, location,

and site. The first level, industry orientation, refers to placement with reference to a source of raw materials, for example logs for sawmills. The second level refers to the location in a particular region that offers most favorable conditions over other similar regions. Finally, site level makes reference to the specific placement of a firm, like a particular city.

Hoover (1948) attempts to formulate principles governing the interrelation of individual locations, the importance of locational change, and how these are relevant to public planning and control. The first factor mentioned by Hoover (1948) that affects the spatial distribution of industries and people is the disposition of natural resources. This is particularly relevant to resource-dependent industries such as agriculture or forestry, and extractive activities such as mining. However, even if all natural resources were distributed evenly over the globe, patterns of specialization and concentration of activities would inevitably appear in response to economic, social, and political principles (Hoover 1948). Hoover stresses that there are certain advantages in concentrating an industry in relatively few locations.

Hoover distinguishes the locational preferences of consumers and producers. According to Hoover (1948) producers' motives are much more significant than consumers' motives in shaping the overall distribution of activities. Geographic differentials in wage rates or the profit prospects of particular occupations are larger and better known than are differentials in living costs or conditions. Producer motives are more compelling as he/she who ignores them risks unemployment or bankruptcy rather than "a mere diminution of the joy of life" (Hoover 1948). Hoover goes on to describe consumers' locational preferences as less tangible as these are strongly shaped by habit and past association. Most people tend to prefer the kind of environment in which they have been living rather than some different social, racial, or institutional atmosphere; an unfamiliar climate and landscape; or change from urban to rural living or vice versa.

Nevertheless, production and consumption are locationally interdependent. Eventually consumer market patterns are determined by the geographic distribution of consumer income, which in turn depends fundamentally on the location of production (Hoover 1948). Locational change is strongly influenced by the extent that production seeks to locate near its markets and at the same time creates market demand, resulting in locational agglomeration.

The producer's choice of location is ultimately driven by the rate of earnings (wages, profits, or interest) attained at different places. Other factors affecting producers' preferences include stability and security of earnings, as well as future prospects of earnings. Locational advantages for producers can be analyzed given the activities incurred by a productive enterprise. Hoover (1948) classifies such activities into three stages: procurement (purchasing and transporting the necessary materials and supplies to the site of processing), processing (transforming the materials into more valuable forms), and distribution (selling and delivering of products). Locational advantages from the standpoint of procurement or distribution (termed by Hoover as transfer operations) depend on the transportation cost of input materials or products. The advantages of sites for processing are governed by production costs which in turn are given by prices of factors of production and available technology that determines amounts of those factors needed per unit of output.

In general, a longer distance involves greater transportation and transfer expenses. However, certain important qualifications must be noted. The distance in question is not measured as a straight air line but rather along the most economical (least-cost) route (Hoover 1948). Topography and climate often determine what routes are easiest to institute and maintain, and hence, have much to do with the variation of transfer costs.

Producers have an incentive to locate as near as possible to their suppliers and markets in order to reduce transfer costs (Hoover 1948). However, transfer costs do not vary simply and

directly with distance. Transfer is canalized along organized routes forming networks. Costs and rates are generally less than proportionately greater for longer hauls on one route. For example transfer costs are lower in the direction of lighter traffic flow, rise discontinuously with increasing distances, lower for large shipments and large shippers and lower for compact and easily handled goods and goods of low value in proportion to weight (Hoover 1948).

Transfer costs shape industry locational pattern. According to Hoover (1948) the ideal location for production processes on the bases of transfer costs from a single input materials source and to a sole market will generally be at the source or the market, rather than anywhere else in between. Orientation closer to the source of materials is based either due to a weight loss in the process or on higher transfer costs per ton-mile on materials than on final products as was also previously suggested by Weber (1929). Orientation closer to the market happens in industries that are involved in the final stages of processing and handling goods that usually involve differentiation, subdivision of consignments into smaller lots, more value in relation to weight, and greater perishability in both physical and style terms. As a wide generalization, Hoover (1948), considers that early stages of production are material-oriented and late stages of production are market-oriented while intermediate stages are relatively foot-loose as to transfer considerations. Flexibility in the combinations of materials used or of products turned out increases the area of locational choice and generally favors orientation to material sources or markets rather than intermediate points.

Regarding the orientation of a whole industry, Weber (1929) suggests that single location of production is the exception and a split of production into several locations will be the rule for productive processes that can technically be divided. Furthermore, splitting of an industry is facilitated when more input materials are used and when these additional materials are used in several independent stages of production. There will be no need to split the production process

on the basis of rates of transportation when only the first manufacturing stage involves the use of several materials and later stages only require additional application of labor. In such a case, the first stage will occur near the location of input materials, while the location of all remaining stages will take place along the way between the location of the first stage and the place of final consumption. The remaining locations will almost always be situated at the place of consumption, due to advantages of the market. Despite the large number of possible independent stages of production, the process will be split into only two stages, which Weber refers to the stage of materials and the stage of consumption.

Isard (1949) defined the general Theory of Location as one embracing the total spatial array of economic activities, with attention paid to the geographic distribution of inputs and outputs and the geographic variations in prices and costs. The classical location theory focused on the analysis of how factor inputs are transformed into physical commodities (McCann and Sheppard 2003). Hence, the spatial dimension is seen as a factor input that determines the characteristics of the transformation process.

Hoover (1948) deems that the understanding of how different factors of production are priced helps determining the geographical distribution of industrial activity. Hoover (1948) groups the needs of a producer in four categories: equipment, a site, labor, and government (i.e. law enforcement). Differentials in the prices of productive services arise mainly from the difficulty or expense of moving factors from one place to another. To the extent that any factor of production is mobile, it moves to places where it is better rewarded. This results in a reduction of price differentials across geographic areas. As a consequence, the mobility of investment funds reduces differentials in interest rates, and the mobility of labor reduces differences in wages. The price of a freely mobile factor would be the same everywhere and would not affect the location of production or other factors at all. Hoover stresses that the magnitude of price

differentials corresponds inversely to mobility. Hence, land shows the largest differences, partly because it is immobile and partly because there is a large variation in the natural endowments of sites.

Rawstron (1958) suggests three principles governing industrial location. The first principle, physical restriction, refers to the fact that choice of location is restricted when a natural resource is the main production input, and hence production is restricted to the availability of such resource. The second principle, economic restriction, stresses the effect on the choice of sites when the cost of one of the inputs to the manufacturing process varies widely from place to place. Rawstron (1958) considers the locational analysis of the cost structure, coupled with the spatial variation in the cost of each input in the cost structure, as a realistic approach to the study of industrial location. Rawstron's third principle is related to technical restrictions and technological change. According to this principle, industries that experience continuous technological changes pay little attention to location factors. However, industries that tend to undergo dramatic changes in technology that require establishing new factories will in fact consider location factors more carefully.

In 1826 von Thünen published his major work "The Isolated State" in which he attempted to explain the laws that govern the prices of agricultural goods and the causes behind patterns of land use (Chisholm 1968). von Thünen considers that if land is rented to the highest bidder those activities that generate the highest rents will be placed at each location. The gradient of the bid-rent curve for each industry is the transport or commute costs to the market located in the center (Kilkenny 1998). With increasing distance from a market there is consistent decline in the ceiling rents payable by any one type of land use. Models for spatial allocation in the von Thünen tradition are useful for studying pattern of industrial use of space and are appropriate for analyzing rural development (Kilkenny 1998).

Von Thünen came to the conclusion that a land use structure similar to concentric rings of different types of cultivation around a central urban market location would develop for the agricultural sector (Figure 4). The level of transportation cost influences the slope of rent gradients. When transport is costly, the top rent for any given kind of use drops off rapidly with increasing distance. However, along a route of cheap transport, the corresponding rent gradient is relatively flat (Hoover 1948). The key conclusion of von Thünen's concentric-rings is that an upward-sloping supply curve for a product can be generated simply by changes in land use (McCann and Sheppard 2003). Higher market prices for a good allow for higher potential rental payments, which motivates the use of larger areas of land to come under cultivation and produce bigger output quantities. Also, von Thünen considered the facility with which a commodity can be transported to the central market (Chisholm 1968). This is incorporated in the analysis of land uses in the form of transportation costs per unit of area from the production to the final consumption site.

The upper diagram in Figure 3 shows the relation between distance from a market and rent in four different types of land uses A, B, C, and D. The lower part of the diagram is a map of the resulting pattern of idealized land-use zones which depicts the concentric rings described by von Thünen. The rent gradient rises to a peak in the market city, since that would be the optimum location for each use from the standpoint of distribution costs alone. The rent gradients fall at different rates, so that each use in turn appears as the highest bidder.

It would be exceptional to find such land use pattern in the real world. One reason it does not occur is the irregularity with which transfer costs correspond to distance. Another reason is that each product or kind of land use has its own geographic pattern of supply areas and market centers (Hoover 1948).

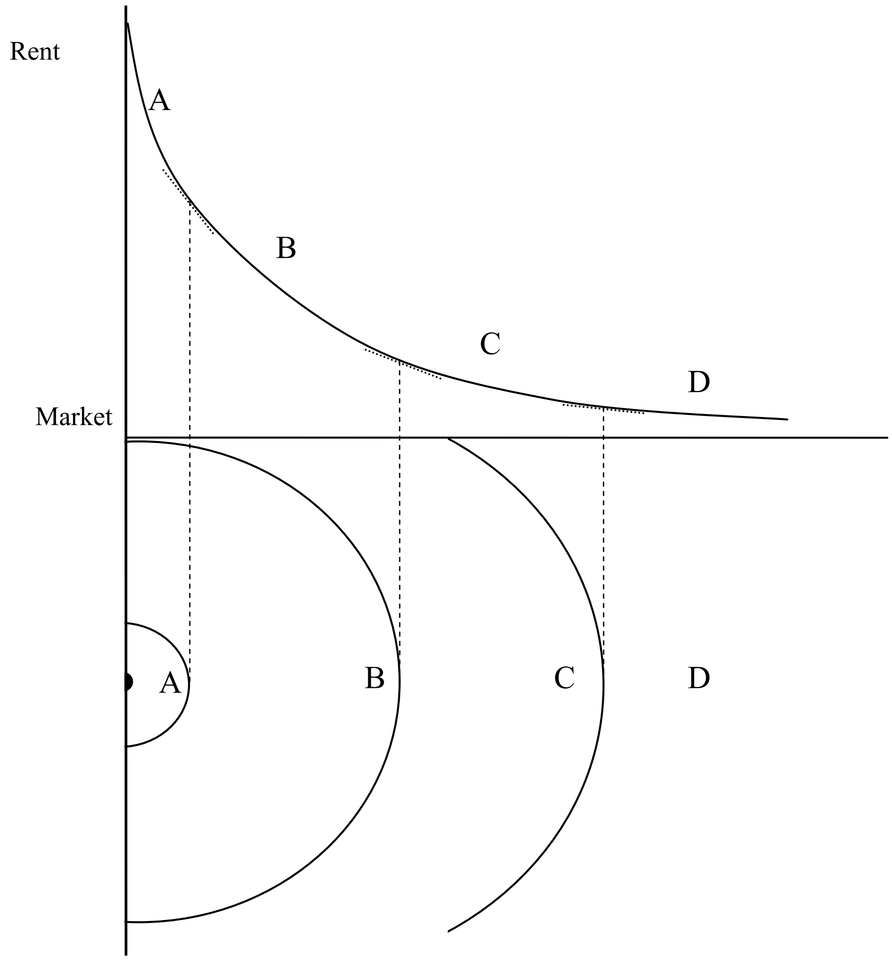


Figure 3. Rent gradients and zones of land use tributary to a single market center in the von Thünen tradition.

Source: Hoover (1948).

In order to minimize processing costs, the individual enterpriser seeks a location conducive to high utilization of the productive capacity of factors and scale of output appropriate to that location (Hoover 1948). The best combination of factors involves more intensive utilization of any factor where its price is high, the most conspicuous variations occurring in the rent of land.

A more recent approach to location theory and land use is presented by Kilkenny (1998). Kilkenny (1998) provides a framework for analysis where $\pi^m{}_i$ denote profits of firms in the i^{th} market-oriented industry located at distance *m* from the market center. Profits are given by total

revenues minus average costs of producing quantity of output Q, minus transportation costs and land rates for the m location. Formally this is given by,

(21) $\pi^m{}_i = P_iQ_i - AC^m{}_i - t_imQ_i - R^m$

where P_i and Q_i represent price and output for industry i, AC is the average cost of production at the m location, t is the output transport-cost rate per unit per mile from the market, and R^m represents land rents at location m. The models assumes that delivered output prices are the same across locations and a free entry equilibrium which zero profit condition results in

(22) $R^m = P_iQ_i - AC^m{}_iQ_i - t_imQ_i$

which illustrates how bid rents are inversely related to average production and transport costs at location m. Holding output Q constant the above expression can be rewritten to

(23) $R^m = (P_i - AC^m{}_i)Q - t_imQ.$

Notice that lower transportation costs result in an increase in distance at which rents could be paid by the i^{th} industry. This relation results in an extension of the distance from the center at which those activities may be located as the transport-cost rate declines. This is depicted in Figure 4.

One reason that firms within an industry locate near one another, noted by Ricardo, is to reduce the cost of transporting inputs. Firms trade lower input-transportation costs for higher output-transportation costs in order to maximize profits. The geographic concentration of production within a country often results in the specialization of regions in one or a few main industries. As a consequence of industrial specialization consumers can benefit from lower cost of outputs. Despite the benefits of specialization these come along with the risk of higher unemployment. Thus, workers who live in industrially more specialized labor markets should be compensated in the form of higher wage rates (Diamond and Simon 1990).

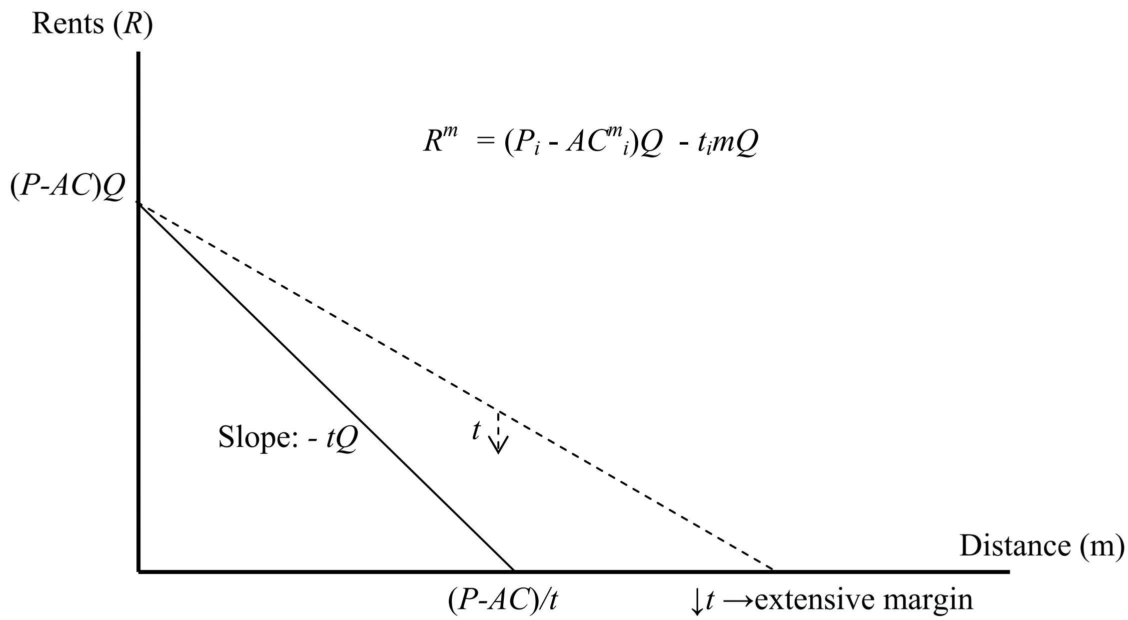


Figure 4. Effect of lower transportation costs on rents and extensive distance margin

3.2 New Economic Geography

Following on the steps of the Theory of Location, Paul Krugman and Masahisa Fujita developed what they have called the New Economic Geography (NEG). According to Fujita and Krugman (2004) NEG attempts to explain the formation of a large variety of economic activities agglomeration in geographical space. The agglomeration or clustering of economic activity occurs at different levels from concentration of small shops in a local neighborhood to a North-South dualism in the global economy.

NEG tells a story that explains both concentration and dispersion of companies. Compared to traditional location theory and economic geography, NEG introduces a general equilibrium model for the entire economy. Furthermore, NEG considers the role of increasing returns to scale at the level of the individual producer or plant that leads to a market structure characterized by imperfect competition. It also considers the role of transport costs and ability of production factors to move in space and allow for industrial agglomeration. Spatial

agglomeration is a result of equilibrium between forces that support concentration (centripetal forces) and those that oppose it (centrifugal forces).

Krugman (1991, 1995) presents a model with two regions, two sectors (a monopolistically competitive manufacturing and a perfectly competitive agriculture) and two types of labor (farmers and workers). This $2 \times 2 \times 2$ model illustrates how interactions among increasing returns to scale at the firm level, transport costs and factor mobility influence spatial economic structure. The model further assumes that the manufacturing sector produces a variety of differentiated products using workers as its solely input. Each product is manufactured by a separate firm experiencing scale economies. The agricultural sector produces a homogenous good under constant returns to scale and uses farmers as the only production factor. Krugman (1991, 1995) further assumes that workers can freely move between regions while farmers are immobile and homogeneously distributed between the regions. Another assumption is that agricultural goods are traded between regions at no cost, while interregional trade of manufactured goods involves a positive transport cost.

Fujita and Krugman (2004) identify the immobility of farmers as the centrifugal force opposing concentration because farmers consume both agricultural and manufactured goods. Centripetal forces are more complex to explain and include a dynamic circular causation. First, a greater variety of goods are produced in a region if a larger number of firms locate in that region. Then, workers, who are also consumers, enjoy a better access to a greater number of product varieties compared to workers in the other region. Besides, workers in the region with a higher firm concentration experience a higher level of income (keeping other things constant) and motivate workers to migrate to the higher paid region. The increase in the number of workers results in the creation of a larger market compared to the other region. Because of scale economies, firms have an incentive to concentrate production in one region and because of

transportation costs it is more profitable to produce in the region offering a larger market and ship to the other, other things equal. Centripetal forces emerge from a circular causation of forward linkages and backward linkages. Forward linkages represent the incentive of workers to be near a concentration of manufacturing production because it will be less expensive to buy goods provided at a central production location. Backward linkages refer to the tendency of manufacture production to concentrate where there is a large market, and a market will be large where manufacture production is concentrated. This relation is what Myrdal (1957) called circular causation (cited in Krugman 1991). Table 8 presents a list of centrifugal and centripetal forces.

Table 8. Forces affecting geographical concentration and dispersion

Centripetal forces	**Centrifugal forces**
Market size effects (linkages)	Immobile factors of production
Large labor markets	Land rents/commuting
Knowledge spillovers	Congestion
Pure external economies	Pure diseconomies

Sources: Fujita and Krugman (2004), Krugman (1998).

If centripetal forces overcome the dispersing effect of centrifugal forces the economy will end up with a core-periphery pattern in which all manufacturing is concentrated in one region. A core-periphery pattern is likely to occur when transportation costs of the manufactures is low enough, when products are sufficiently differentiated, and the expenditure on manufactured products is considerable in the whole economy (Fujita and Krugman 2004).

Ellison and Glaeser (1997) consider two types of agglomerative forces, which they refer to as spillovers and natural advantage. Locational spillovers include gains from sharing labor markets, gains from inter-firm trade, the effect of local knowledge on the location of spin-off firms, and any other forces that might provide increased profits to firms locating near other firms in the same industry. Natural advantages refer to natural endowments that for example have lead

to the development of the wine cluster in California or the agglomeration of large shipyards to locate near bodies of water.

Positive spatial spillovers or agglomeration economies, termed by Ciccone and Hall (1996) as thick market effects, entail that production is more efficient when it is spatially concentrated. Firms benefit from the proximity of other firms within the same industry. However, negative spillovers can also emerge from firm agglomeration resulting in a counteracting effect. Cohen and Paul (2005) consider that the combined effects of positive and negative externalities on an industry cost structure are not obvious and deem empirical investigation to quantify and analyze their patterns.

As it has been mentioned by Krugman (1998) centripetal forces motivating industrial agglomeration emerge due to positive externalities (lower transportation costs, availability of skilled labor and specialized inputs, and knowledge spillovers). But firms face external motivations to avoid densely populated areas because of high competition and high prices for inputs. There are also negative externalities caused by location at a distance from major urban areas such as limiting communication or lower transportation infrastructure. The balancing of positive and negative factors is particularly complex in industries where primary productive inputs accrue in rural areas, but demands concentrate in urban areas as is the case of the food manufacturing industry presented by Cohen and Paul (2005).

3.3 The Theory of Clusters

Ross (1896) was probably the first to mention that industry clustering emerges as a result of economic gains from enterprises interactions. Alfred Marshall (1920) identifies three main forces driving industrial geographical agglomeration/clustering. The first one is the presence of a large labor market pool. The second advantage is the provision of intermediate goods and services. These include raw materials, supplies, consultations and collaboration. The last

component is the occurrence of knowledge exchanges and spillovers between nearby firms and institutions. Marshall (1920) further stresses the importance of externalities of specialized industrial locations to the geographic concentration of companies. Harold Hotelling (1929) suggested that firms will have agglomerative tendencies in a market where buyers of a commodity are uniformly distributed and their purchasing decisions are solely based on the price of the good and transportation costs. Hoover (1948) considers that industries using jointly produced materials or turning out jointly demanded products have an incentive to locate in nearby places. Hoover further expands on the importance of a locally available pool of labor as an inter-industry linkage that involves the use of complementary production factors that promotes agglomeration effects. Hoover (1948) stresses that certain operations and services that an enterprise would have to do for itself in an isolated state, can be instead outsourced to other separate enterprises specializing in those functions and operating in a large enough scale to do them at a lower cost.

But the Theory of Clusters is attributed to Michael E. Porter that developed it during the 1990's. Cluster development theory is premised on the notion that companies tend to spatially concentrate in places where they experience unusual competitive success. A cluster is a critical mass of interconnected companies and associated institutions in a particular field in a particular location, linked by commonalities and complementarities (Schmitz 1995, Porter 1998b). Clusters are geographic concentrations of a group or groups of companies encompassing related industries in an industry supply chain (Porter 1998ab, 2000). They may include input suppliers, ancillary service providers, or providers of specialized infrastructure. Clusters can extend horizontally or vertically to take advantage of production and commercialization efficiencies. An example discussed by Porter (1998c) is the California Wine Cluster. This cluster includes

680 commercial wineries and several thousand independent wine grape growers, along with input and service suppliers as well as local educational institutions that support the industry.

Clusters enhance competitiveness in three ways according to Porter (2003). First, firms can improve productivity because transaction costs are low. Second, clusters foster innovation by increasing the ability of companies to perceive opportunities for new products, new processes, and meeting new needs due to the sheer concentration of entities in the field. Third, clusters facilitate the commercialization of innovation by lowering the barriers to entry of new firms via startups, spin-offs and new business lines of established firms.

Furthermore, the study and identification of clusters can contribute to a better understanding of contemporary industrial patterns, processes of industrial transformation, industry competitiveness, and regional development (Hallencreutz and Lundequist 2003, Peneder 1995). According to Hallencreutz and Lundequist (2003), the current shift in industrial and regional policies towards adopting cluster-based economic development strategies highlights the importance of clustering in current business models.

Porter (1998b) considers that the ability for industries to attain competitive advantage influence the location of firms. According to Porter (2000), previous thinking of the influence of location on industrial competition has relied on rather simplistic views of how companies compete. The dominant view in the post-World War II period rested on endowments of generic factors of production such as natural resources, capital, and labor (Porter 1998a). This view considers competition to be static and mainly based on cost minimization in a relatively closed economy. Comparative advantage in factors of production, resulting in lower production cost, is the key component in such type of analysis. Locating a firm in close proximity to similar types of firms or suppliers/demanders may have economic motivations in terms of enhanced productivity

or lower costs. According to Cohen and Paul (2005) firm agglomeration may occur because of factors such as conglomeration of specialized inputs and information or knowledge spillovers.

Although factor endowments continue to play an important role in locational competition, factors per se have become less valuable as more countries enter the global economy, as national and international markets for inputs become more efficient, and as the factor intensity of competition diminishes. Factor endowments continue to influence the location of industries, particularly resource-dependent and labor-intensive activities, but play a diminishing role in determining wages and standard of living (Porter 1998a).

Porter (2000) stresses that real competition is dynamic. It depends on industrial innovation and on the search for strategic differences rather than pure lower cost advantages. Close linkages with buyers, suppliers, and other institutions are important to a cluster rate of improvement and innovation. Location affects competitive advantage through its influence on productivity and particularly on productivity growth. Productivity and prosperity of a location lie not on the industries in which its firms compete but rather on how they compete. Productivity refers to the value created per day of work and unit of capital or physical resource employed (Porter 1998a). Firms can be more productive in any industry if they employ sophisticated methods, use advanced technology, and offer unique products and services, regardless of the type of industry (e.g. either if it manufactures shoes or semiconductors). All industries can make use of high technology and be knowledge intensive.

The sophistication of how companies compete in a location is influenced by the quality of the microeconomic business environment. Porter (2000) mentions some of the aspects of the business environment such as the road system, corporate tax rates or the legal system. Firms cannot employ advanced logistical techniques unless a high-quality transportation infrastructure is in place, or firms cannot compete using high-service strategies unless they can access well-

educated people (Porter 1998a). These economy-wide areas are key to the emergence of industrial clusters and represent major constraints to competitiveness in developing economies. In more advanced economies, Porter (2000) deem that essential aspects of the business environment for competitiveness are cluster specific such as the presence of particular types of suppliers, skilled workers, or university programs.

The move to an advanced economy requires developing energetic local rivalry. Rivalry must shift from low wages to low total costs, which demands upgrading the efficiency of manufacturing and service delivery (Porter 1998a). Rivalry should also evolve beyond cost to include product and service differentiation. Competition must shift from imitation to innovation and from low investment to high investment, not only in physical assets but in intangibles such as skills and technology.

Porter (1998b) divides the context for strategy and rivalry in two dimensions. One is the climate for investment. A rising investment intensity of competition is necessary to support more sophisticated forms of competition and higher levels of productivity. Macroeconomic and political stability set the context for investment, but microeconomic policies are also important such as the structure of the tax system, the corporate governance system, labor market policies affecting workforce development incentives, and intellectual property rules and their enforcement, among others. The second dimension of the context for competition according to Porter (1998b) is local policies affecting rivalry itself. Local rivalry can be affected by openness to trade and foreign investment, government ownership, licensing rules, antitrust policy, and the influence of corruption.

Porter (1998b, 2000) modeled the effect of location on competition using four interrelated influences: factor (input) conditions, demand conditions, context for firm strategy and rivalry, and related and supporting industries. This model is reproduced in Figure 5. Factor

inputs are comprised of tangible assets, information, the legal system, and university research institutes that firms draw upon in competition. Specialized inputs, in particular those fundamental to innovation and upgrading such as specialized university programs, promote high levels of productivity but tend to be location specific (less available from elsewhere).

Porter (1998b) highlights the importance of macroeconomic and political stability in encouragement investment. Investments allow for the adoption of new technology that facilitates improvements in productivity and cluster specialization. But Porter also stresses the role played by microeconomic policies in a firm's final decision to locate. Microeconomic policies include the structure of the tax system, the corporate governance system, labor market policies, and intellectual property rules and their enforcement, among others.

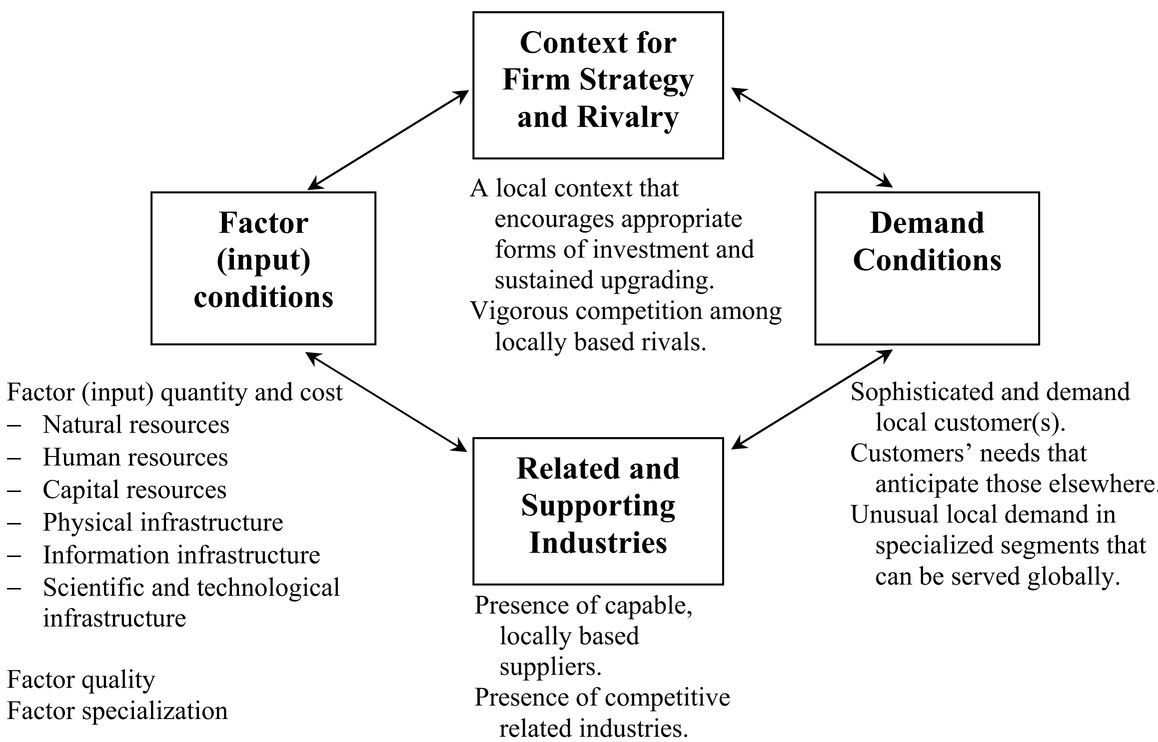


Figure 5. Michael Porter's sources of locational competitive advantage.
Source: Porter (1998b, 2000).

The presence of university research, academic programs and or extension programs influences the emergence of clusters. Knowledge spillover constitutes one of the business environment necessary conditions mentioned by Porter (2000) to develop competitive business clusters. Jaffe (1986) suggests that firms directly benefit from spillovers of academic research and, thus, are motivated to locate near university campuses and research centers. Jaffe (1989) studied the spillover effects from university research to commercial innovation. He used state-level time series data to model the effect of corporate R&D, and university research on the development of corporate patents. His results suggest a positive effect of university research in the Drugs, Electronics and Nuclear technology.

Jaffe et al. (1993) suggest there is empirical evidence that knowledge spillovers are geographically localized using patent citations as a proxy for the generation of new knowledge. Localization fades over time, but at a slow rate. A survey among inventors in the U.S. carried out by Jaffe et al. (2000) also suggests that knowledge spillovers are more likely to occur within a state rather than nationwide. However, Thompson and Fox-Kean (2005) consider that the methods used by Jaffe et al. (2000) that provide evidence for localized knowledge spillovers may include a spurious component. Thompson and Fox-Kean (2005) suggest that the analysis of patent citations as a proxy for knowledge spillover actually provides weaker evidence of its geographic concentration at the state level. Nevertheless, their results still favor a certain level of knowledge spillover agglomeration in space.

Local concentration of any industry fosters the development of a labor force particularly productive in that industry, in the form of skilled and experienced labor. The effect is to strengthen and continue the concentration of firms using similar labor skills. Where the degree of labor skill required is high and at the same time the product is non-standardized, there is the

creation of a double incentive to concentration in few centers for at least that part of the industry in which demands are more particular and the product is more specific.

3.4 Choice Theory and Random Utility

Economic choice theory is based on psychophysical analysis which states that decision-makers have the capacity to discriminate between various stimuli and determine their most preferable option (Thurstone 1927). Lancaster (1966) broke away from the traditional approach that considers that goods are the direct objects of utility. Instead, Lancaster argues that goods per se do not give utility to the consumer but rather the attributes or characteristics of goods give rise to utility. Lancaster (1966) summarizes his approach to consumer theory as one that (1) considers that a good has a particular set of characteristics, characteristics which result in utility, (2) a good possess more than one characteristic and many characteristics are shared by more than one good, and (3) goods in combination may have characteristics different from those pertaining to the goods separately.

Lancaster (1966) imposes several assumptions to develop a model for consumer behavior. First, Lancaster associates a scalar (a_{jk}) with a consumption activity k, for a j good (x_j) in a linear relationship. Second, that a consumption activity (y_k) produces a fixed set of r characteristics (z_r) available to the consumer, keeping a linear relationship with consumption activity and characteristics (b_{rk}). And third, that an individual has an ordinal utility (U_i) as function of the product vector characteristics (z).

These assumptions result in a set of equations

$$x_j = \sum_k a_{jk}\, y_k \quad \text{or} \quad x = \mathbf{A}\, y$$

$$z_r = \sum_r b_{rk}\, y_k \quad \text{or} \quad z = \mathbf{B}\, y$$

$$U_i(z) = f\,(z_1, z_2, \ldots z_r).$$

Here, matrices **A** and **B** capture consumption technology which is an important determinant of consumer behavior. Lancaster further assumes that consumption technology is static to ease the discussion and application of the consumer behavior model.

Louviere et al. (2000) present an analysis of the standard Lancaster approach so that it can be interpreted in terms of discrete-choice models. The decision making process underlying discrete choice model can be represented in the form of the following interconnected equations:

$s_k = f_{kr}(z_r)$, $u_j = g(s_{kj})$, $P_j = h(u_j)$

and

$P_j = h\{ g [f_{kr} (z_r)] \}$,

where s_k is the consumer perceived marginal utility of good k, z_r is an observable value of the objective characteristic r, u_j represents the overall utility associated with the jth alternative, s_{kj} indicates the level of attribute k associated with alternative j, P_j account for the likelihood of choices allocated to alternative j, and f, g and h represent functional forms.

According to the Lancaster's (1966) simplified model, goods x can be transformed into objective attributes, z, in the following relation

$z = \mathbf{B}\, x$,

where **B** is an R by J matrix which converts the J good (set of alternatives in a choice set) into R number of objective characteristics (alternatives, attributes). Utility can then be expressed as a function of the commodity characteristics as follows:

$u = U\,(z_1, z_2, \ldots, z_R)$

where z_r is the level of the r^{th} characteristic that a consumer derives from commodities ($r = 1,\ldots, R$).

Building on this idea of random utility, modern economic choice theory considers that individuals behave in such a way that their preferences are maximized. Preferences may contain

random components due to fluctuations in perceptions, attitudes and other unmeasured factors. Preferences can be defined over goods with complex hedonic attributes, both measured and unmeasured. Choice theory is made operational by linking the random preference model to market response probabilities. An axiomatic structure places response probabilities in statistical models and distributions that allow for coefficient estimation and analysis (McFadden 1986).

Following Rosen (1974), Louviere et al. (2000) present a simplified model where a consumer maximizes utility subject to income constraints such that individuals maximize $U(z_1, z_2, \ldots, z_R)$, subject to $p(z_1, z_2, \ldots, z_R) + d = M$, where $p(z_1, z_2, \ldots, z_R)$ comprises the price of one good with characteristics $z_1, z_2, \ldots, z_R$ which are acquired, d is the price of all other goods and M is the consumer's income. Prices are determined by the distribution of consumer tastes and producer costs, and price is a function of a fixed value of the vector z (Rosen 1974).

The ith decision maker has a utility function defined over a large number of differentiated products (Rosen 1974). Louviere et al. (2000) suggest a conceptual framework that explains the complex decision making and the choice process in six steps (Figure 6). First, a consumer becomes aware of needs and problems that need to be solved. This step is followed by a period of information search during which information is gathered in regard to the types of products that can satisfy the needs or solved the problems identified in step one. Next, consumers evaluate and compare the different alternatives that are available to attain their objectives and any uncertainties associated to these choices. Once consumers become acquitted about a product category a decision rule is formed that maximizes their utility. This utility function considers valuing and trading off product attributes that matter in their final decision. In a fifth step, consumers develop a preference for products that is ordered depending on the product attributes and alternatives. A decision is made whether a product is purchased or not depending on budget

constraints. The final step refers to post-choice (re) evaluation. A general framework for the decision process is presented below.

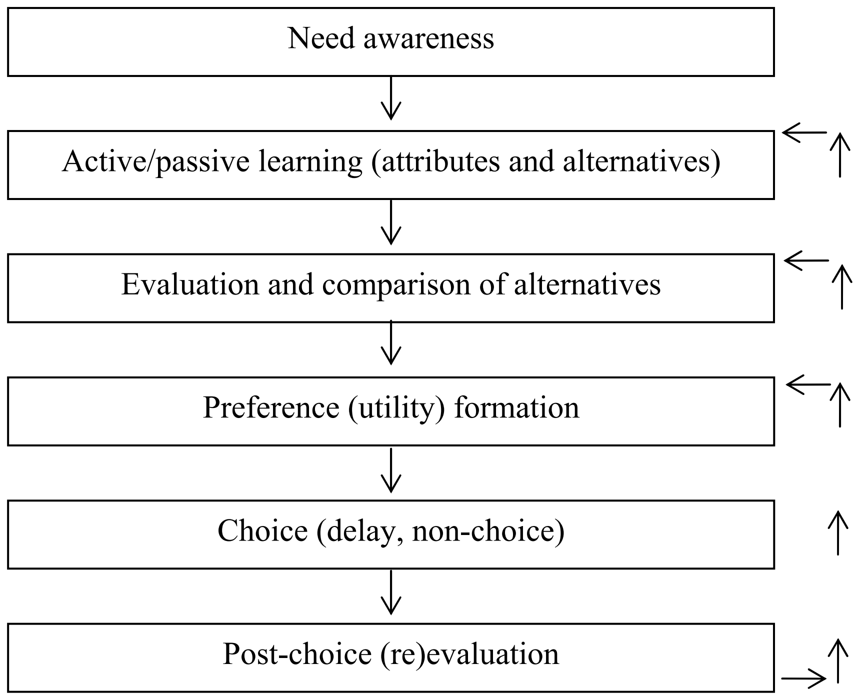


Figure 6. General framework for consumer choice decision process.
Source: Louviere et al. (2000).

According to MacFadden (1974) and Greene (2003) an unordered-choice model can be motivated by a random utility model. A decision maker (*i*) has an utility function defined over an array of *J* possible choices. The decision maker utility function depends over the choices different attributes, which he/she takes as exogenously given, and could also be influenced by his/her own characteristics. A random utility model arises when it is assumed that a consumer utility function contains a deterministic component and a set of unobservable variables that introduce a random error element (Hanemann 1984). This concept combines the ideas of variation in tastes among individuals in a population and unobserved random variables in econometric models.

For the ith consumer faced with J choices, the assumed utility of choice j can be represented by:

$U_{ij} = V_{ij} + \varepsilon_{ij}\,,\ j = 0, \ldots, J$

$V_{ij} = x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots x_{ij}\beta_j.$

The V's represent scale values or strict utilities which summarize the preference of the ith consumer for the jth alternative. Here the x's fully specified functions of measured attributes, characteristics or self-explicated scales of the site aspects, and the β's denote importance weight parameters. Vector $\varepsilon_{ij}, j = 0,1,2,\ldots J$ captures unobservable variables affecting tastes, is stochastic and reflects the idiosyncratic of the individual in tastes for the alternative with attributes x. Here x_{ij} is a $1 \times K$ vector that differs across alternatives and possibly across individuals and is non-stochastic.

The random utility model assumes that a consumer selects the option that maximizes his/her utility. Then, if consumer selects option j among J choices, it is assumed that U_{ij} is the maximum among the J possible utilities (Greene 2003). The statistical model is driven by the probability that choice j is made by the ith individual, which is:

$P_{ij} = \text{Prob}(U_{ij}{*} > U_{ij} \mid V_{ij} = k_j, j \vee \{C_i\}),\ \forall\ j \neq j{*}.$

If the elements of the stochastic error term in the utility function are to be independent from each other and identically distributed across individuals and alternatives then, the probability of decision maker i choosing the jth option is given by the function of the deterministic portion of the decision maker i's utility for site j and a function of the deterministic portions of the same individual utilities for all options in the choice set (Punj and Stalein 1978). These assumptions entail that preferences do not vary across the population (i.e. if vector β represents the weight coefficients for all decision makers and not for a particular ith individual such that β_i is not represented by $\beta + \delta_i$ where δ_i represents an idiosyncratic factor). Furthermore,

the model assumes no omitted variables in the model for V_{ij} which would result in ε_{ij} being correlated to ε_{ik}. If and only if the J disturbances are independent and identically distributed with a cumulative distribution function $F(\varepsilon_{ij}) = \exp(-e^{-\varepsilon_{ij}})$, then the double exponential distribution (Punj and Stalein 1978, McFadden 1986) results in:

$$P_{ij} = \frac{\exp\{Vij^*\}}{\sum_{j=1}^{Ji} \exp\{Vij\}}$$

Which represents the probability of individual i to choose site j. Here site j is one observation of set choice C={1,2,…, J}.

When the odds ratio between P_{ij}* and probability of choosing site j** (P_{ij*}/P_{ij**}), is independent of the presence or absence of a third alternative because $\text{Log}(P_{ij*}/P_{ij**}) = V_{ij*} - V_{ij**}$, then the Independence from Irrelevant Alternatives (IIA) axiom holds because it entails that including another alternative or modifying the characteristics of a third alternative does not affect the relative odds between alternatives j and h.

The IIA axiom has positive and negative implications. The IIA axiom makes it possible to infer choice behavior in multiple alternatives using data from paired comparisons, and makes forecasting the demand for a new alternative an easy procedure. It also allows for simple data processing by permitting to analyze samples of alternatives from large choice sets. Negative implications include the assumption of a uniform pattern of response to changes in the attributes of one alternative. This assumption often conflicts with heterogeneous patterns often found in economic and marketing problems (McFadden 1986). Hausman and McFadden (1984) developed a formal test of the IIA assumption based on the detail that the vector of parameters β can be consistently estimated by conditional logit by focusing on any subset of alternatives if the conditional logit model is true.

The model for random utility U_{ij} considers population heterogeneities that appear as variations in model coefficients across decision makers, or in disturbance variance components that are present across subjects. If the subject effects are considered as unique to the individual, then coefficients must be estimated solely from data on that individual. This model is known as fixed effects model. For estimation, extensive data on a subject is necessary to estimate a fixed effects model accurately, which requires experimental design that tests comprehensively a small number of subjects, and omits consideration of external factors that are consider homogenous within subjects.

When subject effects for individuals drawn from a population are random variables with a probability distribution, instead of a fixed effects model, a random effects model is considered instead. Random effects models can be estimated using sparse data on individuals, with estimation accuracy derived from statistical regularities across a large number of individuals (McFadden 1986). Random effects models capture parsimonious patterns of behavior in the population, and can produce accurate forecasts of the market behavior. Furthermore, random effects allow for pooling of data in which external factors vary, permitting identification of the effects of these factors as a central part of the statistical analysis. However, random effects models cannot make accurate predictions of the behavior of single individuals. In the methods section specifics of the Conjoint Analysis that is based on utility maximization for site selection are presented.

CHAPTER 4. RESEARCH MODEL AND HYPOTHESES

The general research model for the study of the spatial location of a resource-based industry lies on the assumption of an industry with equal cost structure as in Rawstron (1958). The foundation of the research model is based on the traditional theory of the firm where enterprises seek utility maximization through profit maximization. Profit maximization is achieved by producing at the least-cost possible location given the current state of technology. Although managerial theories of the firm as in Williamson (1963), and Baumol (1968) assign priority to some goal other than profit maximization due to managerial objectives, in contrast to the traditional profit-maximizing case, utility maximization is still pursued by the decision makers. Then, a firm should locate in a place where it can maximize utilities to the decision maker and in this research will be treated as equivalent to profit maximization. This component of the study further assumes that decision makers will act just as consumers in an open market, where location is interpreted analogous to a final product that a consumer chooses over a variety of options. The selected product (in this case a given location) is the one that provides the highest level of utility. A given site is characterized by an R number of z characteristics or attributes from which utility is derived, as in the random utility model. It is assumed that a large number of differentiated sites are available so that choice is made among various combinations of z. Further assumptions include full availability of information such that decision makers can identify the site of his/her highest utility given the full array of possible choices, freedom to choose any location and costs of establishment are considered fixed and homogenous across locations.

The study of the behavior of the Softwood Lumber Industry as an example of a resource-based industry, should represent a systematic pattern of preference for site attributes. Based on the work of Ross (1896), Renner (1931), Jaffe (1989), Seldon and Bullard (1992), Bigsby

(1994), Murray (1995), Porter (2003), Aguilar and Vlosky (2006) various factors have been identified as affecting the geographic aggregation of resource-based industries. Among the factors that cause a resource-based industry to set in a particular location we have:

- Sufficient supply of raw materials,
- Nearness to sources of raw or auxiliary materials,
- Sufficient sources of energy,
- Costs of transportation,
- Land values,
- Residence of consumers and location of major markets,
- Availability and low cost of labor (management, skilled workers and unskilled workers),
- Presence of a favorable business environment,
- Occurrence of adequate technical knowledge and inventive talent,
- Knowledge and talent spillovers,
- Environmental legislation.

Figure 7 summarizes these factors. The problem of establishing an industry, therefore resolves itself into assembling several factors upon a selected locus where one of more of them may already occur. In the tradition of Krugman (1991, 1998) and Fujita and Krugman (2004) many of the factors behind spatial aggregation can be classified in Centrifugal and Centripetal forces depending on whether these disfavor or promote geographic concentration.

A three-step research component looking at decision-makers stated preferences and actual industrial behavior is developed. The first component involves the survey of decision makers in the softwood lumber industry that will be asked about the importance of various factors influencing their choice to select a location for a hypothetical new sawmill and those

relevant to the current sawmill location. This first component will be used to determine common factors behind selected attributes and identify the relative importance of each component in the decision-maker choice.

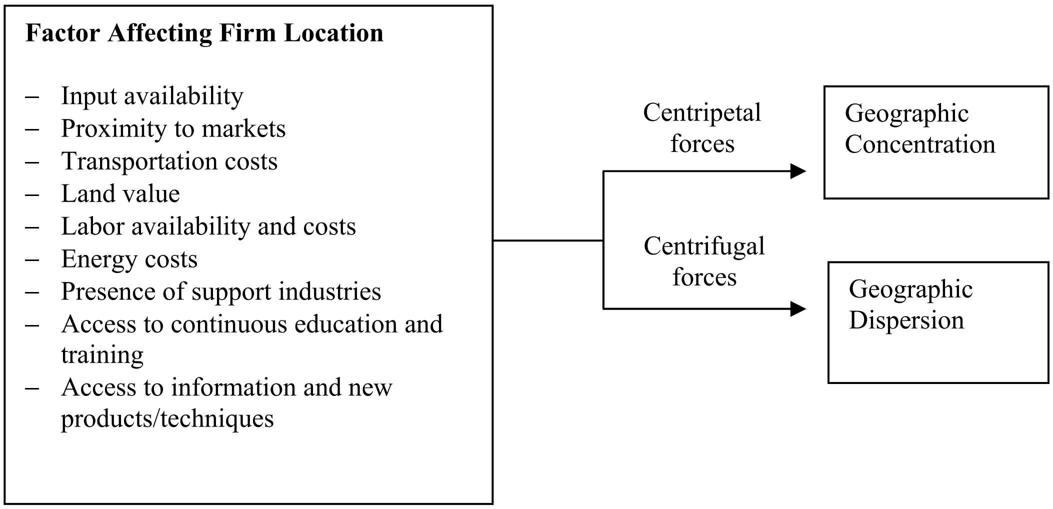


Figure 7. Factors influencing the spatial concentration of companies in the Forest Products industry.

The second component consists of a static, cross sectional analysis that will attempt to detect a cause-relationship effect between selected explanatory variables and the presence of softwood sawmill enterprises. The variables included in the model are those selected as the common factors driving the location of the current sawmill location.

A final stage looks at the evolution of the lumber industry over the 1999- 2005 period. While looking at the behavior of the industry over a decade a growing or comprising trend is identified. These results will be accompanied by respondents' perceptions over the future of the industry and what location factors encourage or discourage the emergence of clusters in the industry. This component captures some of the centrifugal and centripetal forces as suggested by the New Economic Geography and Clusters theory.

While the first component of the research looks at stated preferences in regard to the attributes that managers consider influence their selection of location for a new enterprise, the other two components look at actual firm behavior in the way that the industry aggregates in space, the. Hence, by using a dual approach this research attempts to determine the key elements involved in the spatial location of a resource-based industry.

The presence of softwood lumber enterprises (y) is the dependent variable in the model that is modeled as a function of socio-economic, ecological, transportation attributes, and knowledge spillover variables identified in the second method component. Formally, this can be expressed as:

$$y_{SL} = f(S,E,M,K) \qquad SL = \text{softwood lumber enterprises}$$

where S is a vector of Socio-economic variables, E represents ecological variables, l, M is transportation attributes, K is a proxy for knowledge spillover. The selection of the actual variables will be determined by industry stakeholder preferences (Stage 1). The spatial econometric study of sawmill sector in the U.S. should be able to identify the key components affecting profit maximization to the average sawmills. Based on the literature review a set of hypotheses are formulated for testing using different techniques and approaches presented in the Methods Chapter.

4.1 H_1: The Primary Input Material to the Lumber Industry, Logs, Is the Most Important Factor Determining the Location of the Industry

The price of the inputs, adequate supply and quality of raw materials is the main driver of a resource-based industry. The supply of raw material is essential to the development of the industry as in Ross (1896), Marshall (1920), Hoover (1948) and Porter (2000). Regions with lower prices for logs, the primary input for the industry, are expected to be more attractive and hence capture a higher concentration companies.

4.2 H_2: The Cost of Energy Has a Significant Effect and an Inverse Relation with the Likelihood of Firm Location

Weber (1929) mentions the cost of energy and fuel as a major factor influencing location. Porter (2000) also mentions energy as a factor affecting the geographical aggregation of industries. Formal models will test for the effect and significance of price of energy on the likelihood of sawmill company location.

4.3 H_3: Labor Costs and Availability Have a Significant Effect on the Choice to Locate a New Softwood Lumber Company

Locations with high wage rates do not necessarily attract job seekers or repel employers. The best labor supply from the industry standpoint may be found in places with relatively high wages (Hoover 1948). Porter (2000) also considers that labor productivity is a more important driver than simple wages. Labor availability is a major factor that attracts industry (Marshall 1920, Krugman 1998). Nevertheless, assuming that labor is equally productive in a homogenous region, areas with lower wage rates will be preferred over sites with higher wage rates.

4.4 H_4: Access to Transportation Venues Is a Factor that Has a Significant Effect on Attracting Industry

von Thünen (1826), Weber (1929), Kilkenny (1998) stress the importance of transportation cost to industry location. Transportation costs can be reduced by easy access to main venues that facilitate transportation of goods to final markets. The research question whether the presence of major transportation venues like highways or ports have a significant effect on sawmill location is tested.

4.5 H_5: As a Resource-based Industry the Softwood Lumber Industry Locates Near the Source of Raw Materials

Distance to the primary input is relatively more important than distance to markets (Hoover 1948). Availability of suppliers is a key link to the development of the industry (Porter

2000) and should have a positive and significant effect on the likelihood of sawmill location in a particular region.

4.6 H_6: The Presence of Substantial Final Markets Influences the Location of Softwood Lumber Companies

The presence of substantial final markets is not considered a major driver of the Softwood Lumber Industry location. Aguilar and Vlosky (2006) did not find that a large consumer market drives the location of Primary wood product manufacturers. Instead, Secondary Wood Product manufacturers tend to locate near final markets. Being close to final consumers represent a market advantage to firms as it is easier to identify any changes in consuming preferences or tastes (Kotler and Armstrong 2001). The sawmill sector, as a primary wood product manufacturing, should not find locations with a high market concentration more appealing than remote less populated areas. Tests for the significance effect of this factor should provide evidence favoring this argument.

4.7 H_7: Land Rent Theory, What Is the Effect on Softwood Lumber Enterprises Location?

In the tradition of the von Thünen theory for the agricultural sector land use can develop in a core urban area surrounded by agricultural areas. Peripheral areas characterized for being in areas of lower land rents than their urban counterparts. Hence, the likelihood of finding sawmill companies should be accompanied with lower levels of land rents.

4.8 H_8: The Presence of University Programs and Research Institutions Has a Significant Effect on Softwood Lumber Industry Location

Adam Jaffe (1989) has suggested a positive effect between academic and private research and certain industry sectors. The research question remains whether there is in fact a positive and significant effect of the geographic coincidence of formal academic and research forestry programs with the presence of forest products manufacturers and in particular to sawmill

companies. The questions whether sawmill companies consider the presence of university research programs as an important factor when locating will be explored.

4.9 H_9: New Economic Geography, Do Centrifugal and Centripetal Forces in the Krugman and Fujita Tradition Influence Industry Location?

Several factors that influence the aggregation or dispersion of industries will be tested to determine their behavior in the Krugman and Fujita tradition. It is expected that two different forces, centrifugal and centripetal will be detected in the study.

4.10 H_{10}: Do Preferences for Location Factors Vary Across Decision Makers?

Krugman (1995) argues that one of the reasons why traditional location theory failed to be widely expected is that fact that it did not identify who was the decision-maker and the location decision was analyzed as a firm decision. Harrison and Sambidi (2004) for example, sampled CEOs of major U.S. broiler companies. They consider CEOs to be the decision-makers in the industry and who would maximize their level of utility. This study will include company owners and managers in the analysis of preferences making it possible to test whether they have different factor preferences.

CHAPTER 5. RESEARCH METHODS

The research methods to study the geographic concentration of Softwood Lumber enterprises include three stages. First, data is gathered from self-administered questionnaires through mailed surveys to obtain information on respondent profile, company characteristics, location factor preferences and perceptions on industry structure and future developments in the Softwood Lumber sector. Information is also used to identify centrifugal and centripetal forces affecting firms' location decision. Second, a Spatial Regression Model is developed using Econometric and Geographic Information Systems (GIS) tools to represent the relation between Industry location and explanatory variables. Third, two cross-sections are compared to explore the expansion or contraction of the industry.

This section describes the methods followed to analyze the interdependence of preferences for location attributes using Common Factor Analysis. Common Factor Analysis is used as a tool for data reduction to determine the main factor affecting location in the Softwood Lumber Industry. The relative importance of location preferences is studied using a Conjoint Analysis approach.

Next, the methods to study the relation between explanatory variables derived from the Common Factor Analysis carried out in step one of this research and the presence of the Softwood Lumber industry are described. A general methodology for the study of the evaluation of the industry over a period of time (1999-2005) is presented as part of the analysis.

5.1 Methodology for the Analysis of the Interdependence of Location Attributes Using Common Factor Analysis

Factor analysis is a generic name given to a type of multivariate statistical methods which primary aim is to define the underlying structure in a data matrix (Hair et al. 1998). The two primary applications for factor analysis are summarization and data reduction. Factor analysis

attempts to describe the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors. The factor model is based on the assumption that all variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group (Johnson and Wichern 2002). Factor Analysis can also be used as a tool for developing new empirical typology. Rummel (1970) affirms that Factor Analysis can be used to group interdependent variables into descriptive categories in the basis of similar profile values.

Factor analysis then allows for the understanding of data in a much smaller number of concepts than the original individual variables. Compared to other tools for data analysis that explicitly establish a relation between dependent and independent variables, factor analysis is an interdependence technique in which all variables are simultaneously considered (Hair et al. 1998). In factor analysis, the variates (factors) are formed to maximize their explanation of the entire variable set, not to predict a dependent variable(s). Shook (1999) considers that factor analysis helps to group a large number of attribute variables into a reduced number of uncorrelated and homogenous factors. In this study factor analysis is used to reduce a large number of factors that affect sawmill location. Factors are selected based on a review of the literature on industry location theory and sector specialists. Furthermore, selected factors are used as tool to develop a model for site selection that will aid in the identification of explanatory for the spatial econometric model for the Softwood Lumber industry.

Johnson and Wichern (2002) consider factor analysis to be an extension of principal component analysis as both try to approximate a covariance matrix for the variables in a dataset. The total variance found in a matrix can be broadly divided into several components, but are generally classified into common and unique variances (Rummel 1970). Common variance refers to the variance of a variable X_i that is common to the remaining variables in a matrix of m

variables. Unique variance, or the uniqueness of a variable X_i, refers to the variance component of X_i that is not common to the other m-1 variables in a matrix of m variables. Uniqueness can be further subdivided into specific and random error variance. The specificity of a variable X_i is that portion of its unique variance that is reliable, while the random error is its unreliable unique variance. Thus, the general model assumes a relationship where the value of a random variable is given by:

X_i = Mean + common variance + unique variance (specificity + random error).

The orthogonal factor model can then be described in the following fashion (Johnson and Wichern 2002). An observable random vector, X, with p components, has mean μ and covariance matrix $\mathbf{\Sigma}$. The factor model proposes that X_i is linearly dependent upon a few unobservable random variables $F_1, F_2, \ldots, F_m$, called common factors, and p additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_p$, or error terms. Then, the factor analysis model is given by:

$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \ldots + \ell_{1m}F_m + \varepsilon_1$

$X_2 - \mu_2 = \ell_{21}F_1 + \ell_{22}F_2 + \ldots + \ell_{2m}F_m + \varepsilon_2$

...

$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \ldots + \ell_{pm}F_m + \varepsilon_p$

or, $X - \mu = \mathbf{LF} + \boldsymbol{\varepsilon}$, where x and μ are vectors of dimension p x 1, $\mathbf{L}$ is a matrix ($p \times m$), $\mathbf{F}$ is a vector ($m \times 1$) and $\boldsymbol{\varepsilon}$ is a vector of errors ($p \times 1$). The coefficient ℓ_{ij} represents the loading of the ith variable on the jth factor, such that $\mathbf{L}$ is a matrix of factor loadings. The error ε_i is particularly associated to the ith response Xi. The p deviations $X_1 - \mu_1, X_2 - \mu_2, \ldots X_p - \mu_p$ are expressed in terms of $p + m$ random variables $F_1, F_2, \ldots$, Fm, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ which are latent (unobservable variables). To make it feasible to estimate, the model further assumes that the unobservable random vectors $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ satisfy these conditions:

- **F** and **ε** are independent, Cov $(\mathbf{F}, \boldsymbol{\varepsilon}) = \mathbf{0}_{pxm}$, $E(\mathbf{F}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{F}) = \mathbf{I}$, and

- $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & ... & 0 \\ 0 & 0 & 0 & \psi_p \end{bmatrix}$.

Also, the number of common factors m must be much smaller than the number of variables measured p. The orthogonal factor model implies a covariance structure for **X**. Johnson and Wichern (2002) derive this structure to be

$(\mathbf{X} - \boldsymbol{\mu})\ (\mathbf{X} - \boldsymbol{\mu})' = (\mathbf{LF} + \boldsymbol{\varepsilon})\ (\mathbf{LF} + \boldsymbol{\varepsilon})' = (\mathbf{LF} + \boldsymbol{\varepsilon})\ ((\mathbf{LF})' + \boldsymbol{\varepsilon}') =$

$= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\ \boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$.

So that the covariance matrix $\boldsymbol{\Sigma}$ can be derived from the common factor model

$\boldsymbol{\Sigma} = \text{Cov}\ (\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu})\ (\mathbf{X} - \boldsymbol{\mu})' = \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{LL}' + \boldsymbol{\Psi}$.

The model $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$ is linear in the common factors. The portion of the variance of the ith variable that is given by the m common factors is referred as the ith communality. The portion of $\text{Var}(X_i) = \sigma_{ii}$ due to the specific factor is often called the uniqueness or specific variance. Then, σ_{ii} can be expressed as:

$\sigma_{ii} = \ell_{i1}^2 + \ell_{i2}^2 + ... + \ell_{im}^2 + \psi_i$ or $\ell_{i1}^2 + \ell_{i2}^2 + ... + \ell_{im}^2 = h_i^2$

which results in $\sigma_{ii} = h_i^2 + \psi_i$, $\quad i = 1,2, \ldots, p$

where h_i^2 is the ith communality and ψ_i is the uniqueness or specific variance. Instead of the covariance matrix $\boldsymbol{\Sigma}$, a correlation can be used instead as this is simply the covariance matrix for the standardized variables. The goal of common factor analysis is to determine (a small) common factor dimensions that can reproduce the space of the common parts of the data vectors (Rummel 1970).

Factor analysis can identify the structure of relationships among variables or respondents by studying the correlations between variables or the correlations between respondents. If the

input data matrix is derived from the computation of correlations between variables, then it would be an R-type factor analysis. If the correlation matrix refers to the correlations between individuals, then a Q-type analysis should be used.

The component factor model is appropriate when the primary concern is about prediction or the minimum number of factors needed to account for the maximum portion of the variance explained by the original set of variables (Hair et al. 1998). The number of factors to extract remains a criterion left to the researcher. Hair et al. (1998) suggests setting a predetermined criterion, such as the percentage of variance or latent root criterion used to arrive at a specific number of factors to extract. For example, the percentage of variance criterion approach achieves a pre-specified cumulative percentage of the total variance extracted by successive factors. Hair et al. (1998) consider that in social sciences it is common to consider solutions that account for 60 percent of the total variance to be satisfactory. The latent root criterion, instead, is based on the rationale that any individual factor should account for the variance of at least a single variable if it is to be retained for interpretation. Each variable contributes a value of 1 to the total eigenvalue. Thus, all factors having latent roots or eigenvalues greater than 1 are considered significant while factors with latent roots less than 1 are not significant and may be disregarded. Hair et al. (1998) deem the use of eigenvalues to be reliable to set a benchmark for factor selection when the number of variables is between 20 and 50. Regarding sample size, Hair et al. (1998) recommend not to use factor analysis for samples of fewer than 50 observations.

The factor matrix contains factor loadings for each variable on each factor. The first factor is the single best summary or linear relationships exhibited in the data. The second factor is the second-best linear combination of the variables, subject to the constraint that this is orthogonal to the first factor. To be orthogonal to the first factor, the second factor must be

derived from the variance remaining after the first factor has been extracted. Therefore, the second factor is a linear combination of variables that accounts for the most residual variance after the effect of the first factor has been removed from the data (Hair et al. 1998). Factor loadings are the correlation of each variable and the factor. This matrix of unrotated factors provides a solution that may not offer the most adequate interpretation of the variables under examination. A rotation method aims to achieve a simpler and theoretically appealing factor solution. The most common and simple rotation is orthogonal rotation.

An orthogonal transformation corresponds to a rigid (90-degree angle) rotation of the coordinate axes. Let **T** be an orthogonal matrix, which implies that $\mathbf{TT' = T'T = I}$. Then, the common factor model can be rewritten inserting **TT**' without changing the validity of the relationship

$$\mathbf{X - \mu = LF + \varepsilon = LIF + \varepsilon = LTT'F + \varepsilon = L^*F^* + \varepsilon},$$

where $\mathbf{L^*} = \mathbf{LT}$ **and** $\mathbf{F^*} = \mathbf{T'F}$. Thus, any orthogonal matrix **T** can be used to rotate a given solution to a new solution that equally solves the common factor model equation. However, the rotated solution can be easier to interpret (Johnson and Wichern 2002).

Pakarinen (1999) and Shook (1999) provide examples of the use of factor analysis to study preferences for furniture and firelogs respectively. Both studies made use of principal component factor analysis as a tool to reduce a large number of attributes into an easier to understand subset of characteristics for underlying common factors. Both studies made use of principal component factor analysis and performed an orthogonal rotation using the varimax algorithm. Varimax rotation criterion has, according to Rommer (1970), by consensus become the best function for simple analytic rotation. Pakarinen (1999) and Shook (1999) used a pre-specified number of underlying factors for their modeling of products attributes and also a latent root criterion based on the selection of attributes with eigenvalues larger than 1. These studies

identified their selection benchmarks based on the recommendations set by Hair et al. (1998) for samples of different sizes. Hair et al (1998) suggest that when the sample size is 70 loadings of at least 0.65 could be deemed significant, for a sample size of 85 factor loadings higher than 0.60 are significant.

A total of 23 different characteristics that determine the current sawmill location are presented to study participants. Participants are asked to rank these factor using a 1 to 5 scale (1= Not important at all, 3= Neither Unimportant nor Important, 5=Very important) representing their individual importance on where their mill is located. The list of factors is presented in the survey questionnaire in Appendix 1.

5.2 Methodology for the Study of Location Preferences in the Softwood Lumber Industry in the U.S., a Conjoint Analysis

Conjoint Analysis (CA) refers to any method that estimates the structure of respondents' preferences given their evaluations of a set of hypothetical alternatives specified in terms of different levels of selected attributes (Green and Srinivasan 1978). Hair et al. (1998) describe CA as a multivariate technique used specifically to understand how respondents develop preferences for products or services. The technique is based on the premise that consumers evaluate the value of a real or hypothetical product by combining the separate amount of values provided by different levels of the product attributes and choose the one which gives them the most utility (Green and Srinivasan 1978, Carson et al. 1994, Hair et al., 1998).

CA starts with the consumer's overall judgments about a set of alternatives. The analysis then consists in decomposing the consumer original evaluations into separate and compatible utility scales by which the original global judgments can be reconstituted (Greene and Wind 1975). CA includes all models and techniques that attempt to transform individual responses into estimated parameters. A conjoint methodology takes a decomposition approach in which

respondents react to a set of profile description. A profile includes a total description of attributes and their respective levels. CA helps determining a set of part-worths for the individual attributes that, given some type of composition rule, are more consistent with the respondent's overall preferences. The characteristics of the alternatives that the consumer must choose from are considered to have different dimensions, becoming multi-attributes (Greene and Wind (1975). A multi-attribute object is viewed as a bundle of attributes leading to benefits of differential desirability to individuals (Wilkie and Pessemier 1973). The part-worths identify the relative value (importance) the consumer place on particular attributes. In terms of a basic dependent model, conjoint analysis can be described as

$$Y_1 = f(X_{1,}X_{2,}X_{3,}\ldots, X_N)$$

where Y_1 is a non metric or metric response used a proxy for the level of consumer utility, which is a function of the jth level of the pth attribute. Notice that in this basic model it is assumed that the researcher knows the N attributes composing the total value of the product.

Green and Srinivasan (1978) describe three different types of preference models used in conjoint analysis, a vector, an ideal point, and a part-worth function model. Consider a set of t attributes chosen for the good or product in question. This is given by $p = 1, 2, \ldots, t$.

Next, let y_{jp} represent the level of the pth attribute for the jth stimulus. The vector model of preference is given by the following expression for the preference s_j for the jth stimulus:

$s_j = \sum_{p=1}^{t} w_p \, y_{jp}$,, where the $\{w_p\}$ are the individual's weights for the t attributes.

Quadratic or ideal-point models consider that the preference s_j is negatively related to the squared weighted distance d_j^2 of the location $\{y_{jp}\}$ of the jth stimulus from the individual's point of maximum utility $\{x_p\}$. Formally this can be expressed as $d_j^2 = \sum_{p=1}^{t} w_p \, (y_{jp} - x_p)^2$.

The third type of preference model, the part-worth function, instead can be represented by $s_j = \sum_{p=1}^{t} f_p\,(y_{jp})$, where f_p represents a function indicating the part-worth of different levels of y_{jp} for the pth attribute. The part-worth function model is the one with the greatest flexibility allowing for different shapes for the preference function along each of the attributes. Nevertheless, additional flexibility comes at the cost of having to estimate additional parameters that lower the reliability of the model. If an attribute is categorical then the part-worth function is the only one function that is appropriate for correct estimation.

Green and Srinivasan (1978) stress that features of the vector, ideal point and part-worth models can be incorporated in a mixed model. A more detail explanation of models and uses is provided by Green and Srinivasan (1978). Next, I will describe the steps followed to carry out the analysis. It starts with the characterization of the decision problem when the location factors (analog to product attributes) are selected. This step is followed by attribute level selection, selection of method for data collection, experimental design, elicitation of preferences, questionnaire development, and data analysis.

5.2.1 Characterization of the Decision Problem: Selection of Location Factors (Attributes)

According to Lusk and Norwood (2005) one of the biggest challenges when implementing a CA is the construction and statistical design of the product attributes, or in this problem location factors. For example, the selection of combination of attributes and attributes levels to present to participants in a choice experiment. The selection of attributes and levels should be made through focus groups, literature search, interviews with experts, as the decision problem should be clearly presented as the decision maker understands it.

This stage aims to understand how individuals become aware of the need to make the decision in question, how they define the dimensions of evaluation of the product or service, look for information on alternatives and different attributes, construct their choice sets, and finally, make a decision. The characterization of the decision problem should permit to identify sources of individual heterogeneity such as levels of income, education or different perceptions that can result in potential behavioral differences. The outputs of this first stage in the development of an experimental choice-based study are (Adamowicz et al. 1998):

- Identification of a choice set size and composition
- Selection of relevant attributes
- Naming of potential individual differences and,
- Development of a relevant sampling frame for the study.

5.2.2 Attribute Level Selection

This stage is usually carried out along side with the previous stage. It is critical when selecting the levels for the different attributes not to restrict posterior analyses by a limited range of attributes. Attributes are usually identified based on information obtained from focus groups, literature search, and previous experiences. The immediate step following attribute selection is to assign the levels to each attribute. Attribute levels should be selected to reflect relevant levels of variation in current or future markets of interest.

Seven key attributes were selected including average hourly wages, average prices for delivered logs, electricity, average cost per acre of land, quality of roads, distance to main source for logs and distance to final markets.

Two levels were selected for the average hourly wage. A "high" and "low" wage rates were obtained after calculating a regional average for the states comprising the U.S. South. The standard deviation for these observations was calculated. A 95 percent confidence interval was

estimated by adding/subtracting the standard deviation times a 1.96 factor to the regional average. Information for this calculation was obtained from the most recent Annual Census of Manufactures corresponding to North American Industry Classification System (NAICS) code 3211. Code NAICS 3211 refers to manufactures in the Sawmill and Wood Preservation category. This industry group comprises establishments whose primary production process begins with logs or bolts that are transformed into boards, dimension lumber, beams, timbers, poles, ties, shingles, shakes, siding, and wood chips. It also includes establishments that cut and treat round wood and/or treat wood products made in other establishments to prevent rotting by impregnation with creosote or other chemical compounds are also included in this industry group (U.S. Census Bureau 2005). Table 9 presents data corresponding to total production worker wages, hours and average hourly wages in the U.S. South States including regional and national averages for the year 2003. The levels selected were \$15.61 per hour for "high" wage rate and \$10.77 per hour for a "low" wage rate.

Table 9. Total production worker wages, hours and average hourly wages in the U.S. South including regional and national averages for the year 2003 for NAICS 3211 (manufacturing: sawmills and wood preservation).

States	Wages ($1000)	Hours (1000)	Average wage rates ($/hour)
Alabama	142,537	9,521	14.97
Arkansas	142,708	10,323	13.82
Florida	59,373	4,659	12.74
Georgia	144,286	10,947	13.18
Louisiana	52,459	3,804	13.79
Mississippi	126,454	9,335	13.55
North Carolina	144,609	10,150	14.25
South Carolina	82,316	6,222	13.23
Tennessee	62,539	5,135	12.18
Texas	77,307	7,568	10.21
U.S. South Regional Average	**103,459**	**7,766**	**13.19**

Source: U.S. Census Bureau (2005).

The same procedure was followed to determine the price range levels (dollars per ton) for logs to be included in the survey. State-level FOB delivered prices for pine sawtimber were

obtained from the Journal of Southern Timber Prices published quarterly by Timber Mart-South. The most recent data available for the year 2006 are presented in Table 10. Calculations resulted in a higher price level of $62.25 and a lower price of $46.31 per ton of pine sawtimber delivered at the mill.

Table 10. Average pine sawtimber delivered prices (FOB) in the U.S. South for the first and second quarters of 2006.

	Average delivered prices FOB* Mill for pine sawtimber($/Ton)	
State	1st Quarter	2nd Quarter
Alabama	61.17	59.88
Arkansas	51.50	53.50
Florida	54.66	55.25
Georgia	62.38	58.58
Louisiana	55.25	56.00
Mississippi	53.19	49.63
North Carolina	52.50	46.00
South Carolina	56.85	58.00
Tennessee	52.50	51.25
Texas	53.75	52.75
Virginia	49.60	49.91
U.S. South Regional Average	**54.85**	**53.70**

Source: Timber Mart-South (2006a &b). *(1st point of delivery)

Values for electricity cost and a description for quality of roads were taken from the values used by Harrison and Sambidi (2004). Distance to source for logs and distance to final market were estimated based on values reported in the literature.

All attributes included in the CA and corresponding levels are presented in Table 11.

Table 11. Attributes and corresponding levels used in the conjoint analysis of the location preferences in the softwood lumber industry in the U.S. South.

Attributes	Units	Lower Level	Higher level
Average hourly wage in the region	$/hour	10.50	15.50
Average price for logs	$/ton	46.31	62.25
Electricity cost	cents/kWh	4.50	6.50
Average cost per acre of land	N/A	Low	High
Quality of roads from mill to main market	N/A	Poor	Good
Distance to source for logs	miles	70	30
Distance to final market	miles	90	20

5.2.3 Data Collection Alternatives

Data gathering for a CA have commonly involved the use of three methods: (1) a trade-off or two-factor-at-a-time procedure, (2) a full-profile approach, and (3) pair-wise comparison methods (Hair et al. 1998).

The two-factor-at-a-time procedure, also known as trade-off presentation, compares attributes on a two-at-a-time basis by ranking all combinations of levels (Hair et al. 1998). Study participants are asked to rank the various combinations of each pair of attribute levels from most preferred to least preferred. This method is relatively simple to apply and places a low level of information overload on the part of the respondent. However it poses several limitations. It sacrifices realism, the total number of required judgments can be fairly large for even a small number of levels, there is the possibility that participants follow routine-types response patterns, there is no possibility of using pictorial or other non-written stimuli, it allows the sole use of non-metric responses, and it does not permit for the use of fractional factorial stimuli designs to reduce the number of comparisons made (Green and Srinivasan 1978, Hair et al. 1998).

The full-profile approach uses a complete set of factors to describe a product profile. The major risk of this method is the possibility of information overload and as a result, respondents may simplify the experimental task by ignoring variations in the less important factors or by simplifying the factor levels themselves. To avoid information overload a full-profile usually includes no more than five or six attributes. If a greater number of factors are required, then bridging factors should be incorporated in the models. This approach is explained by Green and Srinivasan (1978) and consists in preparing several card decks in which the full set of factors is split into subsets of attributes. One or two factors however remain common among all subsets which are then used as linking part-worth functions. Responses in the two-factor-at-a-time

procedure are measured using ranked orders while the full-profile approach can make use of either rank orders or ratings (i.e. Likert scales ranging from Least Liked to Most Liked).

Green and Srinivasan (1978) consider that the most favorable characteristic of the full-profile approach is that it presents a more realistic description of stimuli by defining the levels of each attributes and taking into consideration potential correlations between factors. But, this method has the disadvantage of presenting the participant with a difficult task as he/she has to consider various factors and levels at the same time.

The pair-wise combination presentation method combines the previous two methods. The pair-wise combination compares two profiles at the time, and the respondent indicates preference for one profile over the other. According to Hair et al. (1998), the distinguishing characteristic of the pair-wise comparison is that the profile typically does not contain all the attributes, as does the full-profile methods, but instead only a few attributes at a time are selected in constructing profiles. It is similar to the trade-off method in that pairs are evaluated, but in the case of the trade-off method the pairs being evaluated are attributes, whereas in the pair-wise comparison methods the pairs are profiles with multiple attributes. The advantages of a choice-based approach using a pair-wise combination are the additional realism and the ability to estimate interaction terms, which are not possible with traditional conjoint analysis (Hair et al. 1998).

Individuals being part of a CA experiment are presented with a variety of choice sets. Each choice set is comprised of different competing options and participants are asked to select the one most preferred option, rank all of them or rate each choice. Every option is characterized by a set of attributes. Thus, the study of choice behavior is described by (1) the objects of choice and sets of alternatives available to decision-makers, (2) the observed attributes of decision-makers, and (3) the model of individual choice and behavior and distribution of behavior patterns

in the population (McFadden 1974). The data observed is assumed to be generated by the trial of drawing an individual randomly from the population and recording her demographics, the set of alternatives available to her, and her actual choice, rating or raking preferences.

5.2.4 Experimental Design

The next stage is to use some form of orthogonal design to generate different combinations, commonly called "profiles", of different attribute levels (Adamowicz et al. 1998). A profile is a single attribute level combination in a complete factorial combination of attribute levels. A design is a sample of profiles which meet particular statistical properties that determines the utility specifications that can be estimated. Frequently, linear model design theory is used to develop stated choices designs.

According to Adamowicz et al. (1998) the majority of choice-based experiments make use of orthogonal arrays commonly known as main effects plans. The design of a choice-based experiment requires the maximization of orthogonality and balance. For example, consider a product with attributes A, B and C, each of which varies at two levels. The concept of perfect orthogonality requires that attributes A, B and C to be uncorrelated with one another. A balance design demands that each level of each attribute occurs with equal frequency. So, for example for two levels of attribute A: A1 and A2 each level should appear 50% of the time in the design. In a balance choice-based design each attribute has equivalent statistical power and the selected attributes are uncorrelated with the model intercept.

The Bretton-Clark designer program was used to select the fractional designs for the study. This program produces a subset of hypothetical profiles based on the attribute levels provided by the researcher. The program minimizes the confounding of attribute main effects by selecting a sub-sample of orthogonal product combinations (Harrison and Sambidi 2004). Nevertheless, Adamowicz et al. (1998) stress the importance of not limiting this stage to canned

designs as they often cannot reflect the research needs or may present unrealistic scenarios. For this reason, profiles generated by the Bretton-Clark designer program were checked for congruency and pre-tested with a focus group.

The eight different profiles generated by the Bretton-Clark designer program are presented in Table 12 based on the attributes and levels previously selected. These profiles are used to estimate a model for site preference and determine the relative importance of each attribute.

Table 12. Location characteristics for the eight profiles generated by the Bretton-Clark designer program

Attributes	1	2	3	4	5	6	7	8
Average hourly wage in the region ($/hour)	15.50	15.50	15.50	15.50	10.50	10.50	10.50	10.50
Average price for logs ($/ton)	62.25	46.31	62.25	46.31	62.25	46.31	62.25	46.31
Electricity cost (cents/kWh)	6.50	4.50	6.50	4.50	4.50	6.50	6.50	4.50
Average cost ($) per acre of land	High	Low	Low	High	High	Low	Low	High
Quality of roads from mill to main market	Good	Good	Poor	Poor	Good	Good	Poor	Poor
Distance to source for logs (miles)	70	30	30	30	30	70	70	30
Distance to final market (miles)	90	90	20	90	20	20	90	90

A challenge for effective market forecasts based on conjoint data is that of omitted variables in experimental design and in reduction of conjoint data using a choice model. When items under study have a large number of attribute dimensions, of which only a small number can be characterized and varied experimentally the participant's imputation of missing variables introduces noise in the system, and potentially bias (McFadden 1986). Omitted variables are a concern in the representation of preferences by a choice model if there are many measure attribute dimensions and levels. The number of possible interactions will often exceed experiment sample sizes, thus, omitting the total number of choices presented to the respondent.

The assumptions underlying the error term in a choice-based analysis reflect the complexity and richness of the choice process by recognizing that a model of this process seldom

will be fully specified in terms that can be measured accurately and which identify all of the current and historical attributes that really influence the choice process. In reality, most models of choice are underspecified and this fact should be taken into account in the analysis (Gensch and Recker 1979). Nevertheless, Wilkie and Pessemier (1973) stress that it is reasonable to expect that only few attributes will dominate a model's predictive and explanatory power in multi-attribute modeling. As long as a model incorporates the key factors affecting decision-making, and it can explain a large portion of the variance in the responses, it can be deemed to be valid.

5.2.5 Eliciting Preferences

Elrod et al. (1992) compared the use of rating/rank-based and choice-based conjoint models based on their predictive ability. Elrod et al. (1992) in a study of housing preferences among graduate students found that there is little reason to prefer the ratings-based or choice-based conjoint approach on the basis of predictive ability. The two approaches predicted equally well on average. The choice of approach may depend more on intended use.

According to Elrod et al. (1992) choice-based models that are fit at the aggregate level offer several advantages. The values and statistical significance of all parameters are easily reported, share predictions for new brands are easily produced. Asking respondents to indicate choices from realistic sets of alternatives closely mimics the market problem. However, aggregate choice models deter segmentation studies of a market. The major challenges to the use of conjoint data according to McFadden (1986) involve (1) the design of laboratory techniques to elicit responses with reliable information on market behavior, (2) the development of methods for converting experimental data into market forecasts, and (3) providing consistent validation of the results.

Two of the most frequently used methods for coding preferences are rank order and interval rating (Harrison and Sambidi 2004). These methods differ in the restriction that each places on the metric and non-metric properties of the subject's preference function. The rank order method demands unambiguous responses from respondents who must rank all hypothetical choices. This provides a non-metric ordering of respondent preferences. Contrary, the interval rating scale methods permits subjects to state order, indifference, and intensity across product choices, a characteristic that allows for both metric and non-metric properties to be elicited. Some information is lost then when ranked order scaling is used because it does not allow subjects to express indifference of intensity across product attributes. Thus, rank order scaling fails to capture cardinal properties in their preference ranking (Harrison and Sambidi 2004).

Ratings-based models are appropriate to segmentation studies, but according to Adamowicz et al. (1998), estimation results are difficult to summarize, tests of statistical significance and simulation of choice shares are cumbersome. MacFadden (1986) mentions that ratings and rankings provide more information on preferences per respondent than choices do. It is a choice of the individual investigator to identify which method is the most appropriate based on empirical grounds.

For this research two forms of preference are used: interval ratings and choice-based model. As mentioned previously, this part of the study will treat the decision making process of a manager to select a location to be akin to consumer purchase decision. Bruner et al. (2001) suggest the use of a seven-point cognitive purchase involvement scale to measure the degree which a consumer's involvement with a purchase is related to utilitarian motives rather than affective motives.

5.2.6 Questionnaire Development

A questionnaire presents a set of profiles and often asks for socio-demographic, psychographic, attitudinal and past behavior data. Past behavior data can provide important information in regard to what a subject did in the past but also any other alternatives he/she may have considered.

As in any other type of survey based research pre-testing of the questionnaire is completed. In a stated choice design the researcher has to define how many choice scenarios each participant will be asked to do. Although there are not set rules in this respect, there should be a balance between respondent learning and fatigue against efficient use of the respondent (Adamowicz et al. 1998). In practice, respondents are usually presented with about eight choice scenarios (Carson et al. 1994). But, depending on the familiarity of the participants with the problem options can be as many as thirty two.

According to Lusk and Norwood (2005) there is a tradeoff between a large experimental design (long questionnaires) and the difficulty of administering it. A large experimental design with good statistical properties can be difficult to administer as it involves the development of several different survey versions or blocks. A methodologically-sound experiment may involve the use or a large number of repeated choice questions but this may call into question the reliability of resulting data. Swait and Adamowicz (2001) report that efficiency of responses to choice-based experiment questions can be significantly affected by the length and difficulty of the choice tasks. Adamowicz et al. (1998) suggest that a large number of choice scenarios can be divided into smaller subsets. Scenarios can be randomized, and then be subdivided to obtain blocks of smaller, more desirable, size. Lusk and Norwood (2005) recommend that researchers should give considerable attention to the simplicity of the survey administration and to easing cognitive burden of survey respondents.

The final version of the questionnaire is presented in Appendix 1. The survey questionnaire is comprised of five sections. In addition to the sections for the interval rating and choice selection of a profile of location characteristics (section 3), it gathers information on the background of the respondent, preferences for softwood sawmill site location factors (used for factor analysis), general information about the mill, and views regarding clustering in the Softwood Lumber Industry.

5.2.7 Data Analysis

Choice-based methods most commonly use limited dependent variable models to establish a relationship between respondent preferences and selected attributes. When respondents are asked to choose between two options, such as either accept or reject a choice with particular attributes, binary response models ought to be applied. A binary response model looks at the probability that a product with characteristics $x_1, x_2, \ldots x_k$ is selected by the respondent. This relationship is given by (Wooldridge 2002):

$$P(y=1|x_1, x_2, \ldots x_k) = P(y=1|\, x) = G(\beta_0 + x'\beta)$$

where G is a function that takes on values between zero and one, β_0 is an overall intercept, and the set of parameters β reflects the impact of changes in x on the probability of y=1. The two most common nonlinear functions used such that G takes on values strictly between zero and one are the logistic and the standard normal (Greene 2003).

When the logistic function is applied it results in the logit model given by

$$G(\beta_0 + x\beta) = G(z) = \exp(z) / [1 + \exp(z)]$$

which is a function that takes values between zero and one for all real numbers z. This is the cumulative distribution for a standard logistic random variable. Instead, the Probit model results

from the application of the standard normal cumulative distribution function which is expressed

as : $G(z) = \Phi(z) \equiv \int_{-\infty}^{z} ø(\upsilon)\mathrm{d}\upsilon$,

where $ø(z)$ denotes the standard normal density: $ø(z) = (2\pi)^{-½} \exp(-z^2/2)$.

Binary choice models interpret the outcome of a discrete choice as a reflection of an underlying regression. This is commonly known as an unobserved latent variable model which is denoted by:

$y^* = \beta_0 + x'\beta + \varepsilon, \quad y = 1[y^* > 0]$

where ε is independent of x and it is further assumed that ε follows a standard logistic distribution or the standard normal distribution (Wooldridge 2002).

In a random utility model framework let U^{A} and U^{B} represent the individual's utility derived from choice A and B, respectively. The observed choice between A and B reveals which one offers the greater utility, but it does not reveal the utility for the other option. Thus, the dependent variable y is equal to 1 if $U^{\mathrm{A}}>U^{\mathrm{B}}$ and 0 if $U^{\mathrm{A}} \leq U^{\mathrm{B}}$. A common formulation for the random utility (latent) variable is given by (adapted from Greene 2003):

$U^{\mathrm{A}} = \beta_0 + x\beta^{A'} + \varepsilon^{\mathrm{A}}$ and $U^{\mathrm{B}} = \beta_0 + x\beta^{B}{}' + \varepsilon^{\mathrm{B}}$.

If preference for A is denoted by $y = 1$, then,

$\mathrm{P}(y{=}1|\, x) = \mathrm{P}[U^{\mathrm{A}}>U^{\mathrm{B}}] = \mathrm{P}[\beta_0{}^{A} + x'\beta^{A} + \varepsilon^{\mathrm{A}} - \beta_0{}^{B} - x'\beta^{B} - \varepsilon^{\mathrm{B}} > 0|\, x]$

$= \mathrm{P}[\beta_0{}^{A} + x'\beta^{A} + \varepsilon^{\mathrm{A}} - \beta_0{}^{B} - x'\beta^{B} - \varepsilon^{\mathrm{B}} > 0|\, x] = \mathrm{P}[(\beta_0{}^{A} - \beta_0{}^{B}) + x'(\beta^{A} - \beta^{B}) + (\varepsilon^{\mathrm{A}} - \varepsilon^{\mathrm{B}}) > 0|\, x]$

$= \mathrm{P}[\beta_0 + x'\beta > 0|\, x]$,

and this probability function is again expressed in the form of $G(\beta_0 + x'\beta)$.

Interval-rating scales are frequently used as a method for coding respondent preferences. The interval-rating scales allow for the expression of ordering, indifference, and intensity across different choices (Harrison et al. 2005). These properties allow for the use of metric and non

metric models for analysis. Ordered responses model is a commonly used method for coding respondent preferences (Harrison and Mclennon, 2004). Rated or ranked order consists of a multinomial response in which values assigned to each outcome are no longer arbitrary (Wooldridge 2002). Ordered Probit and Logit models are used to estimate respondents' preferences for different site attributes. It is also important to consider in the analysis the bounding nature of interval-ratings such as Likert scales. Because of lower and upper bounds the Two-limit Tobit (TLT) model is theoretically appealing as it corrects for censoring and also maintains metric information between bounds.

Let U be an ordered response eliciting a consumer level of utility derived from locating a sawmill at a particular site. Variable U takes on values 0,1,2,3,4,5,6,7 according to the individual level of utility derived from choosing a particular site to locate a new sawmill. The Ordered Probit model for U can be derived from a latent variable model where U^* is determined by:

$$U_n^* = \beta\mathbf{X} + \varepsilon_n \qquad \varepsilon \mid \mathbf{X} \sim \text{Normal}\ (0,1)$$

Where U_n^* is a latent variable representing the nth individual's utility for purchasing a certified product over a non-certified one, β is a vector of part-worth utility effects and the effects associated with selected explanatory variables, **X** is a matrix containing respondents information for the selected variables, and ε is a normally distributed random error term. The ordered response model for this study assumes the following relationship:

$U=1 \quad if\ U^*_i \leq \mu_1;$ $\qquad U=2 \quad if\ \mu_1 < U^*_i \leq \mu_2;$ $\qquad U=3 \quad if\ \mu_2 < U^*_i \leq \mu_3;$

$U=4 \quad if\ \mu_3 < U^*_i \leq \mu_4;$ $\qquad U=5 \quad if\ \mu_4 < U^*_i \leq \mu_5;$ $\qquad U=6 \quad if\ \mu_5 < U^*_i \leq \mu_6;$

$U=7 \quad if\ \mu_6 \leq U^*_{i.}$

Where U is the nth respondent's rating for a particular profile and the μs are unknown thresholds parameters. The values for the dependent variable in the model correspond to the

different levels of attractiveness of a given site profile as denoted by the respondent's stated preference in a 1 to 7 Likert-scale.

The parameters in the model can be estimated by maximum likelihood. For each j (1,2,3,4, 5,6,7), the log-likelihood function is:

$$\ln L(\boldsymbol{\mu}, \beta) = 1[y_j = 1] \ln[\Phi(\mu_1 - \mathbf{X}_i \beta) + 1[y_j = 2] \ln[\Phi(\mu_2 - \mathbf{X}_j \beta) - \Phi(\mu_1 - \mathbf{X}_j \beta)] + \cdots$$

$$+1[y_j = 7] \ln[\Phi(\mu_6 - \mathbf{X}_j \beta)$$

where Φ denotes the standard normal probability function. Replacing Φ with the logarithmic distribution gives the Ordered Logit model. Regression coefficients from the Ordered Probit and Ordered Logit models do not possess the intuitive appeal of normal linear regression coefficients. The marginal probabilities are the expected change in the probability of a particular choice being selected with respect to a one-unit change in an independent variable (Dorfman 1996). To test for marginal effects of variables of interest the response probabilities P(U=i|**X**) were estimated as follows (Wooldridge 2002):

$$\partial p_o(\mathbf{X}) / \partial X_k = -\beta_k ø(\alpha_1 - \mathbf{X}\beta), \; \partial p_j(\mathbf{X}) / \partial X_k = -\beta_k ø(\alpha_j - \mathbf{X}\beta)$$

$$\partial p_j(\mathbf{X}) / \partial X_k = -\beta_k [ø(\alpha_{j-1} - \mathbf{X}\beta) - ø(\alpha_j - \mathbf{X}\beta)], \quad 0<j<J,$$

where again $ø$ denotes the standard normal cumulative density function. The partial derivative of the consumer's utility with respect to a site attribute, $\partial U_n(s^*)/\partial s_g$ gives the value that the nth consumer assigns to the gth attribute (Sy et al. 1997). Specific values for the independent variables considered in the study have to be used when making comparisons across different models to estimate marginal probabilities (Wooldridge 2002).

The TLT model can be specified as follows. Let U_n be the nth respondent unobservable utility derived from a particular combination of location characteristics (attributes) as previously defined. The TLT specification models the observable response rate y (1 to 7) in terms of the unobservable utility as:

$U_n = \beta\mathbf{X} + \boldsymbol{\varepsilon}_n$, and

$y = 1$ if $U_n \leq 1$;

$1 < y < 7$ if $1 < U_n < 7$; and;

$y = 7$ if $U_n \geq 7$

where β is a row vector of part-worth utility effects for the nth respondent and $\mathbf{X}$ is a column vector of location attribute levels (-1, 1) and $\boldsymbol{\varepsilon}_n$ is an error term. Parameter estimates are estimated using maximum likelihood methods as described in Wooldridge (2002).

Among the wide variety of competing choice models in the consumer behavior literature, the Probit/Logit models belong to the general class of models described as covariance models. Covariance models attempt to derive product attribute relative importance that is later used for prediction purposes (Arnold et al. 1981). The method used to elicit preferences determines the model that is more appropriate for the estimation of part-worth values. The Probit/Logit models are consistent with the theory of sampling from a population of utility maximizing decision makers, and often uses attribute ratings of both chosen and not-chosen alternatives in the choice set in order to reveal the determinant attributes. The underlying assumption of the model is that a decision maker can rank, or rate, possible alternatives in order of preference and will always choose an option which is considered to provided the highest level of utility given relevant constraints (Punj and Staelin 1978). Within this theoretical framework Gensch and Recker (1979) consider the Probit/Logit model to be two of the best suitable techniques for cross-sectional multi-attribute modeling.

When an order scale is used the dependent variable is of an ordinal nature and thus, ordered regression models such as Ordered Probit or Ordered Logit are the preferred choice for estimation of part-worth parameters. When interval-rating is used during the elicitation process instead, the best suited model is less clear. Harrison and Sambidi (2004) indicate that TLT

models are commonly used in conjoint analyses such as in Harrison et al. (2002) or Roe et al. (1996). TLT models assume that utility is cardinal (the interval rate is continuous) between upper and lower bounds of the scale. Other studies instead have made use of Ordered Probit or Ordered Logit models (MacKenzie 1993, Sy et al. 1997) suggesting that these are better suited for estimation given that interval rate scales are measured as a discrete variable. However the use of Ordered Probit or Ordered Logit models has the pitfall of assuming that preferences are purely ordinal which fails to capture for cardinal information if respondents express intensity in their responses. Furthermore, ordered models require the use of additional degrees of freedom to estimate part-worth estimates making it less appealing in practice when sample size is small (Harrison and Sambidi 2004). In an analysis of cardinal and ordinal assumptions in conjoint analysis Harrison et al. (2005) conclude that while modern economic theory rejects the equal-interval cardinality assumption and favors the use of Ordered Probit or Ordered Logit over TLT models, application in empirical research provides evidence that parameter estimates from conjoint analyses are not significantly different between the models. The final decision over which model is preferred may depend on whether there are too few degrees of freedom that limit the use of ordered Probit or Logit models.

Linear models have also been used in the analysis of conjoint data. Under this model levels of utility are assumed to be measured directly and the analysis becomes a main-effects analysis of variance (Kuhfeld 1993). The independent variables in the model correspond to the attributes and the development variables the participants' judgments. For example the preference for one individual for a product with three attributes is represented in a linear model by:

$$y_{ijk} = \mu + \beta_{1i} + \beta_{2j} + \beta_{3k} + \varepsilon_{ijk}$$

where y_{ijk} is one subject's stated preference for a product with attributes at the *i*th, *j*th, and *k*th levels. The model assumes that that $\Sigma\,\beta_{1i} = \Sigma\,\beta_{2j} = \Sigma\,\beta_{3k} = 0$. Linear models are seldom used in

the most recent Conjoint Analysis literature. Table 13 summarizes some key characteristics of several model specifications used for Conjoint Analysis.

Table 13. Key characteristics of selected models used in conjoint analysis

Model	Nature of Dependent variable	Measurement scale	Key characteristics
Choice-based preferences			
Logit	Binomial	Binary	Assumes IIA
Probit	Binomial	Binary	Assumes
Conditional Logit	Multinomial	Binary	Can test IIA (Hausman test)
Ranked/rated preferences			
OPM	Choice Ranking/ Interval Rating	Ordinal (Likert Scale)	Uses up more degrees of freedom
OLM	Choice Ranking/ Interval Rating	Ordinal Preferences	
TLT	Interval Rating	Cardinal Preferences	Fewer
LM	Choice Ranking/ Interval Rating	Cardinal/Ordinal	Constant marginal effects (linear relationship)

OPM: Ordered Probit Model, OLM: Ordered Logit Model, TLT: Two-limit Tobit Model, LM: Linear Model.

Table 14 summarizes all variables included in the CA including variable name, type and part-worth coefficients expected signs. All site attributes are represented as binary variables with 1 and -1 values. Variables with higher values are entered in the dataset as a 1 and lower levels with a -1. Notice that all expected signs are negative as higher levels, indicating higher wages, costs or distances are associated with lower preference for that location. The only variable with a positive expected sign is ROAD_QUALITY and relates to the assumption that locations with better quality roads are preferred to those with lower quality roads.

Table 14. Description of variables included in the conjoint analysis of site attributes.

Attributes	Variable name	Type	Expected sign
Dependent variable			
Location profile	RATE	7-point Likert rating scale	NA
Explanatory variables (site attributes)			
Average hourly wage in the region	WAGES	Categorical	-
Average price for logs	LOGS_COST	Categorical	-
Electricity cost	ELECTRICITY	Categorical	-
Average cost per acre of land	LAND_COST	Categorical	-
Quality of roads from mill to main market	ROAD_QUALITY	Categorical	+
Distance to source for logs	DISTANCE_LOGS	Categorical	-
Distance to final market	DISTANCE_MKT	Categorical	-

Following estimation of the model coefficients, the relative importance of each site characteristic is calculated. Attributes' relative importance is estimated using a proxy for the respondent utility following Halbrendt et al. (1991). This proxy is calculated by taking the absolute difference between the highest and lowest values for the *i*th attribute. Then, the relative importance of each attribute is given by:

$$RI_i = R_i / \sum_{i=1}^{g} R_i * 100,$$

were RI_i is the relative importance of the *i*th attribute, and R_i is the proxy of the utility derived from the *i*th attribute.

5.3 Survey Implementation

Survey development and implementation follow methods and procedures recommended by Dillman (1978, 2000) and described as the Tailored Design Method. Data collection was done using a mail survey questionnaire. Mail questionnaires were chosen as the most cost-effective method of data collection. The method affords a high degree of anonymity, is less limited by rigid time constraints that can impede the effectiveness of other research methods and has proven effective when surveying the wood products sector in the U.S. (Vlosky and Chance 2001, Vlosky and Wu 2001, Vlosky et al., 2002). According to the Tailored Design Method, mail questionnaire procedures included survey pre-testing, pre-survey notification of the initial mailing, a post-survey reminder, and a second survey mailing. Table 15 presents the timetable for the different steps followed as part of this study. Non-response bias was evaluated by comparing mean responses received from the initial mailing to those returned in a second mailing (Armstrong and Overton, 1977).

The names and addresses for the participants in the study were obtained from the "Big Book". This directory constitutes the most comprehensive database of buyers and sellers in the

forest products industry. It is produced annually and it is compiled by Random Lengths. Data from the Big Book was gathered for the U.S. South, and the U.S. West (including the West Inland region). Individual Softwood Sawmill information was collected for the year 2005.

Table 15. Mail questionnaire procedures and respective dates for the Choice-based experiment.

Study procedure	Date
Pre-survey notification	October 9, 2006
First survey mailing	October 16, 2006
Post-survey reminder	October 23, 2006
Second mailing	November 13, 2006
Deadline for returned questionnaires	January 31, 2007

5.4 Methodology for the Spatial Analysis of the Behavior of a Resource-based Industry

There is a diverse nature of methods used for the analysis of georeferenced data. For example, Henig and MacDonald (2002) studied the locational decisions of charter schools using an Ordered Probit model. Henig and MacDonald first looked at the occurrence of charter schools at the census tract level in the District of Columbia. To model the incidence of charter schools, tracts with no such schools received a value of 0, tracts with one charter school received a value of 1; and those with two or more received a value of 2. Henig and MacDonald (2002) limited the upper bound of their dependent variable because only a few census tracts had three or more charter schools. Given the ordinal nature of the dependent variable they made use of an ordered Probit model. Blackman et al. (2006) make use of a Probit model to analyze the likelihood that a land plot has been deforested when studying shade-grown coffee areas in El Salvador, Central America (Blackman et al. 2006).

In the following subsections I describe the various steps followed as part of the research project to model the presence of Softwood Lumber companies in the U.S. South. I concentrate on this region because is the one with the highest probability of hosting new developments as expressed by respondents and due to data availability. Methods can be broadly divided into three steps. First, data is gathered from different sources to assemble the database that incorporates

information on the dependent variables in the regression model. Second, once data has been assembled in a common platform, exploratory data analysis is performed to determine the validity of methods that formally incorporate a spatial dimension. Third, formal model development and hypothesis testing is performed using different model specifications.

Methods rely on the use of different software packages depending on the nature of the operation to follow. ArcGIS was the Geographic Information System used in all steps requiring the use of georeferenced information. Exploratory study of the data relies on tools available in S-Plus and ArcGIS. Formal analysis for the different test-statistics and regressions are performed using Stata and the Spatial Econometrics Toolbox developed by James P. LeSage at the University of Toledo.

5.4.1 Assemble of Georeferenced Database

The first step in assembling a georeferenced database is to identify data availability and its sources. A formal spatial econometric model is used as an analysis of industrial behavior which is guided by the results of the Common Factor Analysis performed on the respondents' stated preferences for location attributes.

The first decision when performing spatial econometric analysis is to define the level at which the analysis is performed. The unit sample of analysis is a county/parish. This is done because most of the information used in the analysis is available at the county level, and the use of Federal Information Processing Standards (FIPS) codes allows for the georeference of data with no geographic information by simple code matching. FIPS codes are a standardized set of numeric or alphabetic codes issued by the National Institute of Standards and Technology to ensure uniform identification of geographic entities through all federal government agencies. County boundary lines are obtained from the National Transportation Atlas Database at (www.bts.gov/publications/national_transportation_atlas_database).

Data for the spatial analysis comes mainly in two different formats Shapefiles and Comma Separated Value (CSV) tables. A Shapefile is a file format that allows for the storing of geometric location and associated attribute information. Shapefiles with different attribute information can be merged using the UNION function available in ArcGIS. A CSV file is a form of spreadsheet file that allows for the ease storage of large quantities of data. One of the attributes in the CSV table is the identification of FIPS codes that serves as the linkage between the two file formats.

Data on the number and specific location of sawmills, including latitude and longitude coordinates, is obtained from the U.S.D.A. Forest Service Southern Station (2005). This shapefile is a point file where every sawmill enterprise in the region is identified on a map. Socio-economic data is obtained from Profiles of America, a database released by the Economic Research Service (ERS), the Census Bureau, the Bureau of Economic Analysis and the Bureau of Labor Statistics.

The presence of university research, forest product academic programs may also influence the occurrence of softwood lumber manufacturers. Jaffe (1986) suggests that firms directly benefit from spillovers of academic research and, thus, are motivated to locate near university campuses and research centers. Occurrence of formal forestry/forest products academic programs (i.e. technical and college degrees) are added to the GIS database a proxy for knowledge spillover effects. Table 16 lists institutions in the U.S. South that offer degrees in forestry and/or wood science as listed in the Society of American Forestry (2006) and the Society of Wood Science and Technology (2006) online directories.

Information on forest products sales and number of woodland farms is obtained from the 2002 Agricultural Census (U.S.D.A. National Agricultural Statistics Service 2004). Information on presence of ports, highways, and precipitation are gathered from the U.S. Department of

Transportation (1998) and the NationalAtlas.Gov (2004). These variables were identified a priori but its inclusion in the spatial econometric model depends on the results obtained from the Common Factor Analysis.

Table 16. Institutions offering programs of study in forestry and/or wood science and technology in the U.S. Pacific Northwest and U.S. South regions.

Institution	State	Forestry	Wood Science
Alabama A&M University	AL	Yes	No
Auburn University	AL	Yes	Yes
University of Arkansas	AR	Yes	No
University of Florida, School of Forest Resources and Conservation	FL	Yes	No
University of Georgia, Warnell School of Forest Resources	GA	Yes	Yes
University of Kentucky, Department of Forestry	KY	Yes	Yes
Louisiana State University, School of Renewable Natural Resources	LA	Yes	Yes
Louisiana Tech University, School of Forestry	LA	Yes	No
Mississippi State University, College of Forest Resources	MS	Yes	Yes
North Carolina State University	NC	Yes	Yes
Duke University, Nicholas School of the Environment	NC	Yes	No
Oklahoma State University	OK	Yes	No
Clemson University	SC	Yes	Yes
University of Tennessee, Department of Forestry, Wildlife and Fisheries	TN	Yes	Yes
Texas A & M University, Department of Forest Science	TX	Yes	No
Virginia Polytechnic Institute and State University, Dept. of Forestry	VA	Yes	Yes

5.4.2 Explore Deviations from Complete Spatial Randomness as First-hand Evidence of Industry Clustering

To identify any clustering patterns in the Softwood Lumber Industry in the U.S. South, various inter-point distance methods are used. These include measures of intensity and the empirical distribution for the origin to nearest neighbor point distances. Intensity analysis is carried out using a Binning non-parametric smoothing as a tool for intensity estimation. This method uses a two-dimensional rectangular histogram to form rectangular bins. The counts in these bins are smoothed using a smoothing algorithm (Kaluzny et al. 1997). Deviation of the empirical pattern of sawmill companies will be tested against complete spatial randomness using F-hat, G-hat, K-hat and, L-hat analyses.

F-hat is an origin-to-point nearest neighbors statistic that overlays a $k \times k$ grid on the region under study. Then, it compares the distances from the m resulting origins to their nearest

neighbors. In this analysis an excess of high distance values is indicative of clustering (Kaluzny et al. 1998). G-hat analysis can be used as a tool to determine the clustering of data points. If there is evidence of clustering in the data an excess of short neighbor distances is expected (opposite to the F-hat analysis). An excess of long distance neighbors is evidence of regularity in the data.

The K-hat statistic is used to describe how the interaction between points varies through space. A theoretical value for this statistic under a spatial point process with no spatial dependence allows comparing it against the empirical K-hat statistic. Deviation from complete spatial randomness provides evidence in favor of clustering. The L-hat function under spatial homogeneity closely resembles a straight 45 degree line. Deviations of the empirical statistic from complete spatial randomness simulations suggest the presence of point clustering.

A Pearson's Chi-square test for deviation from CSR assuming a Poisson distribution is also evaluated. Under the assumption of stationarity, observations follow a spatial process where the overall mean value is a constant (Cressie 1993). A Chi-square test-statistic can be calculated using the mean value of the count-data by county as an estimator for the overall mean. Deviations from a Chi-square distribution will provide evidence of different geographic company frequencies across the state.

5.4.3 Tests for Spatial Dependency

The detection of spatial autocorrelation among regression residuals implies either a non-linear relationship between the dependent and independent variables, the omission of one or more spatially correlated explanatory variables, or the appropriateness of an autoregressive error structure (Florax and de Graaff 2004). It is important to apply a test for spatial correlation because ignoring the presence of spatial autocorrelation among the population errors causes

Ordinary Least Squares to generate a biased variance and an inefficient regression coefficient estimators.

There are several exploratory tests to detect the presence of spatial dependency which include the following:

- Moran's IR,
- Lagrange Multiplier principle,
- Kelejian and Robinson (*KR*) Robust approach.

In all three types of tests, the null hypothesis is the absence of spatial dependence (Anselin and Hudak 1992) but they typically differ in the specification of the alternative hypothesis (Florax and de Graaff 2004). Moran's I is known as a "diffuse test", because the alternative hypothesis merely implies spatial autocorrelation among a residual data series. The underlying causes for autocorrelation are unclear. Diffused differ from Focused tests because the latter set a specific alternative hypothesis and usually take the form of a Lagrange Multiplier test (Florax and de Graaff 2004).

Florax and de Graaff (2004) illustrate four different types of spatial dependence tests in the context of an autoregressive moving average spatial process given by:

$y = \xi Wy + X\beta + \varepsilon$

$\varepsilon = \lambda W\mu + \mu,$

$\mu \sim \mathrm{N}(0, \sigma^2 I).$

These are:

- Unidirectional tests, which test either:

Ho: $\xi = 0$ under the assumption that $\lambda = 0$, or

Ho: $\lambda = 0$ under the assumption that $\xi = 0$.

- Multidirectional tests which test for

Ho: $\xi = 0$ and $\lambda = 0$.

- Robust tests, which test for

Ho: $\xi = 0$ under the assumption that $\lambda \neq 0$, or

Ho: $\lambda = 0$ under the assumption that $\xi \neq 0$, which can be assessed on the basis of OLS estimation of the simple linear model without spatial effects.

- Sequential unidirectional tests, which test

Ho: $\xi = 0$ under the assumption that $\lambda \neq 0$, or

Ho: $\lambda = 0$ under the assumption that $\xi \neq 0$, which can be assessed by means of Maximum Likelihood or Instrumental Variables estimation of a specification where one of the spatial parameters is set unequal to zero.

According to Florax and de Graaff (2004) the Kelejian-Robinson (*KR*) test has lower power than Moran's *I*. However, Moran's *I* is not designed to have power against heteroskedasticity which the *KR* test is, Moran's I is not uniformly more powerful than the *KR* test. The power of the *KR* and Moran's I tests depends also on the nature of the data generating process, whether it is an autoregressive or moving average process. For a complete discussion see Florax and de Graaf (2004).

5.4.4 Regional Model for the Occurrence of the Softwood Lumber Industry at the County Level

The use of GIS and statistical regression techniques is a relatively new approach to model natural resource problems, theories of interacting agents and interdependent decision making (Anselin 2000). The presence of the Softwood Lumber industry in a county is estimated using a Geostatistical regression approach. A Geostatistical approach considers spatial variation to be a continuous process yielding a surface of spatial observations (Anselin 2001). The model is

estimated using a correlated errors model and an autoregressive process to adjust for Least Squares estimates.

The occurrence of Softwood Lumber Industry in a county is modeled as a binary variable that takes on values of 0 (no enterprises) or 1 (one or more enterprises in the county). The model can be based on the presence of a latent non-observable variable. In this case the latent variable represents the utility (*U*) derived from placing a sawmill in a given location. Theory states that the decision-maker makes a marginal benefit cost calculation based on the utilities achieved by making the decision to locate in a given county or somewhere else. Then, a model for binary dependent variables derives the latent variable *U* from an underlying latent variable model. Thus, let U^* be an unobserved, variable denoting Utility be expressed as a function of a set of explanatory variables that, in this model, come from the study of common factors from owners and managers' preferences for location attributes. Thus:

$$U_i^* = \mathbf{X}_i\beta + u_{\mathrm{i}} \ (i=1, \ldots n)$$

Where U_i^* is assumed continuous, $\mathbf{X}i$ is a k-vector of exogenous variables, and u_i is a vector of random errors N(0,1). We cannot observe the net benefits of the decision to locate a sawmill in the *i*th county, only whether there is a presence of the softwood lumber industry or not. Therefore, as presented in Greene (2003), dependent variables take on values:

$U_i = 1$ if $U_i^* > \mathrm{T}$

$U_i = 0$ if $U_i^* \leq \mathrm{T}$

where T is an unobservable utility threshold level that determines whether a county exceed or not the utility level identified by the decision-maker as the minimum level to invest in the building or continual running of a sawmill. In a binary response model, interest lies in the response probability that the response is equal to 1, conditional on a set of explanatory variables. This probability is modeled as a nonlinear function *G* that takes on values strictly between zero and

one such that $P(U=1|\mathbf{X}) = G(\mathbf{X}\beta) = G(\mathbf{z})$, and $0 < G(\mathbf{z}) < 1$. In the Probit model G is the standard normal cumulative distribution:

$G(\mathbf{z}) \equiv \int_{-\infty}^{z} \phi(z)\,\mathrm{d}(z)$, where $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ which corresponds to the standard normal density.

Because of the spatial nature of the data used in the model, it is reasonable to assume that nearby observations are correlated. If the exploratory study of data point clustering provides strong evidence of deviation from complete spatial randomness, this result can be used as indicative that the occurrence of sawmills occurs in an aggregated pattern (thus, favoring spatial correlation). Such spatial autocorrelation is commonly modeled by means of a spatial contiguity matrix ($\mathbf{W}$) as in McMillen (1992). Spatial autocorrelation can be expressed as a spatially dependent error model or an autoregressive spatial model. In the case of the former, the model takes the form of:

$$U = \rho\mathbf{W}U + \mathbf{X}\beta + u,$$

or in the later case the error term takes the form

$$U = \mathbf{X}\beta + \lambda\mathbf{W}u + \varepsilon$$

where ρ and λ are the parameters denoting the strength of the spatial correlation and $\mathbf{W}$ is a spatial contiguity matrix. As presented in Anselin (1998) the spatial autoregressive model implies that:

$$U = (\mathbf{I} - \rho\mathbf{W})^{-1} (\mathbf{X}\beta + u).$$

Analogously, the spatially dependent error model implies that:

$$U = \mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1} \varepsilon.$$

Model estimation for the probit model specification can be done using maximum likelihood estimation (Greene 2003). However, as described by McMillen (1992) the extension

of maximum likelihood methods in the presence of spatial correlation is extremely difficult because autocorrelation patterns produce a likelihood function involving numerous integrals, making direct estimation virtually impossible. McMillen (1992) developed a method for estimation for binary response spatial data relying on an EM Algorithm. The name EM comes from the two alternating steps involved in the algorithm that find the expectation (E) of the functions and then maximizes (M) the resulting posterior density to estimate the parameters (Gelman et al. 2004). McMillen(1992) estimation method takes into account spatial correlation and heteroskedasticity (because data vary spatially it is expected that variances vary as well), explicitly. Heteroskedasticity can be a mild source of error in models of continuous dependent variables but it is a serious problem in a discrete dependent variable model. Maximum likelihood estimators are not consistent in the presence of heteroskedasticity (Greene 2003).

As an alternative to the use of the EM algorithm, LeSage (2000) proposes the use of a Bayesian estimation using a Gibbs sampling approach. LeSage (2000) builds on the work of Albert and Chib (1993) on the use of Gibbs sampling for the analysis of binary response data. LeSage (2000) considers that a Gibbs sampling method is superior to the EM approach because it allows for inferences regarding the mean and dispersion of all parameters including the spatial lag parameter that the EM approach cannot. Also, The Gibbs sampling approach produces estimates of the heterogeneous variance for every observation in space and allows prior knowledge to be introduced in the prior distribution. Gelman et al. (2003) provide a detailed explanation of prior and posterior distribution conditional on observed data in Bayesian data analysis

Figure 8 presents a diagram that summarizes the methods used in the study of stated preferences for location factor and the development of a model for industrial behavior that builds on the study participants' responses.

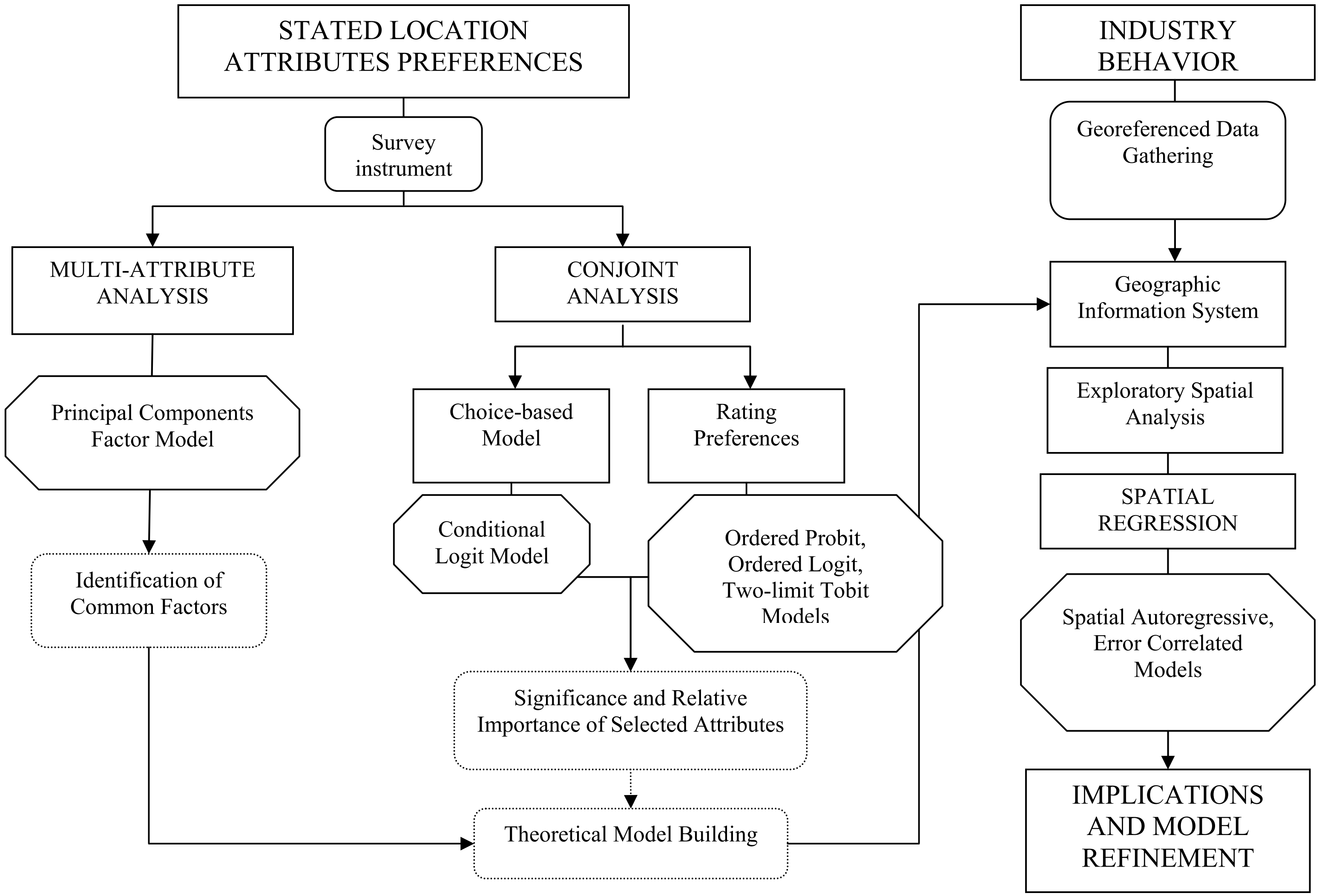


Figure 8. Research methods for the study of location preferences for a resource-based industry

5.5 Analysis of the Aggregation/Disaggregation of Softwood Lumber Enterprises in the U.S. South

Hoover (1948) classifies the basic types of locational changes as seasonal, cyclical, secular and structural according to their character and duration. Seasonal changes are mostly limited to mobile labor and have traditionally been linked to agricultural activities that change from season to season. Cyclical changes last longer than seasonal ones. Hoover (1948) considers that cyclical changes are likely to occur at the country level and affect all economic activities at about the same time, whereas there is considerable diversity in the case of seasonal patterns. Cyclical changes are fluctuations in the total rate of investment with accompanying effects on the total demand for production factors or final products, in particular durable goods. Secular changes or trends are gradual alterations that last for long periods and show no tendency to reverse or repeat themselves as cycles and seasons do. The depletion of an exhaustible resource with use is a good example. Technical progress may also be considered to have a secular trend that results in increased efficiency. Structural changes transform conditions or production techniques and result in a major change in the industry.

Locational shift then occurs as a result of cyclical fluctuations in investment, income distribution, factor utilization, and relative prices. Locational changes also as a consequence of gradual growth of population and depletion of exhaustible resources or the discovery of new resources or developments of new techniques. The shift of wood manufacturing to the U.S. South because of new environmental legislation on public lands in the Pacific Northwest constitutes an example of industrial change because of new policies and regulations. The character and amount of the locational shift that occurs depends on the mobility, adaptability, and elasticity of supply of all the factors concerned (Hoover 1948). Migration of industries from one region to another represents primarily geographic differentials in growth rates rather than

physical transferences of production factors to new locations. Branch plants are more easily involved in deliberate relocations than independent plants are.

Members of the Softwood Lumber industry that participate in the survey of the sector are asked on their expectation in regard to where they would see an increase in softwood sawmill capacity in the next five years. The four major wood product regions in the U.S., the South, West, Northeast and North Central are considered in the analysis to identify in which of them growth is expected to occur in the next five years.

Two cross sections of data on the number of sawmills in the U.S. South and Texas are used to look at the absolute change in the number of operating sawmills in the regions. Although this fails to capture chances in sawmill capacity it can be used to identify whether the industry is contracting or expanding in space and whether there is consolidation in the industry concentrating manufacturing capacity in fewer companies. To further study the effect of forces influencing this process, a list of centrifugal and centripetal forces as presented in the New Economic Geography are presented to survey participants to elicit their perception on the effect in favor or against industry clustering.

CHAPTER 6. RESULTS: MULTIVARIATE AND CONJOINT ANALYSES

Of the total 490 companies listed in the Random Lengths Big Book (Random Lengths 2006), 23 lacked a mailing address. Those companies were called up to request a mailing address and inform them of the undergoing study. After a round of calls to these 23 companies the final database for the study comprised a total of 472 firms.

Of the 472 surveys mailed, 21 were undeliverable because the company had moved or had gone out of business, eight do not manufacture softwood, and five companies requested removal from the study. Of the remaining companies, 81 surveys were returned and usable, resulting in an adjusted response rate of 19 percent. This rate falls between the ranges of recent surveys in the wood products sector such as in Vlosky and Shupe (2004a) 25 percent, Vlosky and Shupe (2004b) 10 percent, Vlosky et al. (2002) 31 percent or Vlosky and Ozanne (1998) 23 percent. Complete questionnaires were received from 21 states (Table 17).

Table 17. umber of usable returned surveys by State.

State	Frequency	Percent	Cumulative Percent
AL	8	9.9	9.9
AR	2	2.5	12.3
AZ	1	1.2	13.6
CA	6	7.4	21.0
CO	4	4.9	25.9
FL	1	1.2	27.2
GA	2	2.5	29.6
ID	3	3.7	33.3
LA	2	2.5	35.8
MS	7	8.6	44.4
MT	4	4.9	49.4
NC	2	2.5	51.9
OK	1	1.2	53.1
OR	11	13.6	66.7
SC	5	6.2	72.8
SD	1	1.2	74.1
TX	4	4.9	79.0
UT	1	1.2	80.2
VA	6	7.4	87.7
WA	7	8.6	96.3
WY	3	3.7	100.0
Total	**81**	**100.0**	

Of the 81 returns 40 correspond to companies located in the U.S. South while the remaining 41 responses came from companies in the Western region. Table 18 summarizes the total number and percentages of the categorical variables included in the survey. The majority of respondents includes plant managers, owners, or holds a position other than the categories included in the questionnaire. In terms of company size, based on number of full-time employees, every size category is represented in the study and is fairly well distributed with 40.7 percent of respondents employing 74 or fewer employees.

Table 18. Summary statistics for the softwood lumber companies included in the study (categorical variables).

Respondents position	**n**	**Percent**	**Cumulative Percent**
Owner	21	25.9	25.9
Sales manager	2	2.5	28.4
Marketing manager	0	0.0	0.0
Plant manager	38	46.9	75.3
Other	20	24.7	100.0
Full-time employees			
No valid	1	1.2	1.2
5 or less	2	2.5	3.7
6-10	3	3.7	7.4
10-24	9	11.1	18.5
25-49	11	13.6	32.1
50-74	7	8.6	40.7
75-99	12	14.8	55.6
100-149	24	29.6	85.2
150 or more	12	14.8	100.0
2005 sales revenues			
Less than $10 million	22	27.2	27.2
$10 – $19.9 million	11	13.6	40.7
$20 – $29.9 million	11	13.6	54.3
$30 – $39.9 million	12	14.8	69.1
$40 – $49.9 million	9	11.1	80.2
$50 – $59.9 million	6	7.4	87.7
$60 – $69.9 million	4	4.9	92.6
$70 – $79.9 million	1	1.2	93.8
$80 – $89.9 million	1	1.2	95.1
$90 – $99.9 million	1	1.2	96.3
$100 – $109.9 million	2	2.5	98.8
$110 million or more	1	1.2	100.0
Finished lumber sold as FOB Mill	75	92.6	
Finished lumber sold as FOB Delivered	63	77.8	

In terms of company revenues respondents represent sawmills in every category. About half of sawmills (54.3 percent) manufacture lumber that in 2005 amounted to $29.9 million or less in 2005. The average company represented in the study has been in operation for 42.1 years, had an annual production of 64.57 MMBF in 2005, procures logs at a maximum distance of 158.05 miles and ships its finished products to places located in an average 1,181.5 miles. When companies are divided into the two regions under study a more interesting picture appears (Tables 19 and 20). Companies in the U.S. South have been in operation for a shorter period of time and have a slightly higher annual production compared to those in the U.S. West. Differences in terms of maximum distance logs are procured and final products are shipped are more significant in absolute terms.

Table 19. Summary statistics for sawmills included in the study (continuous variables) in the U.S. South only.

Description	N	Minimum	Maximum	Mean	Std. Deviation
Years of operation	40	4	100	37.48	24.95
2005 annual production (MBF)	40	2.0	260	73.703	64.895
Maximum distance logs are procured (miles)	37	25	400	114.32	67.12
Maximum distance that finished lumber is shipped	30	100	3000	816.33	830.83

Table 20. Summary statistics for sawmills included in the study (continuous variables) in the U.S. West Region only.

Description	N	Minimum	Maximum	Mean	Std. Deviation
Years of operation	41	8	116	46.63	25.83
2005 annual production (MBF)	41	0.4	226	55.658	51.755
Maximum distance logs are procured (miles)	40	15	800	198.50	167.97
Maximum distance that finished lumber is shipped (miles)	28	90	3750	1572.86	1154.97

To determine if there are statistically significant differences between the softwood sawmills companies in the U.S. South and West regions, t-statistics were computed to determine whether the differences in the mean values of years of operation, annual production, maximum

distances logs are procured and finished lumber is shipped is significantly different from ‘0’. Table 21 shows the results of the tests. At α=0.05 level of significance, softwood sawmills companies in the U.S. South in average procure logs and ship finished lumber to shorter distances than their counterparts in the U.S. West.

Table 21. T-tests for equality of means between companies in the U.S. West and South Region for selected variables.

	Mean Difference	df	Std. Error Difference	t	P-value (2-tailed)
Years of operation	9.16	79	5.644	1.623	0.109
2005 annual production (MBF)	-18.045	79	13.0256	-1.385	0.170
Maximum distance logs are procured (miles)	84.18	75	29.594	2.844	0.006
Maximum distance that finished lumber is shipped	756.52	56	262.852	2.878	0.006

6.1 Multivariate Analysis of Site Location Factors in the Softwood Lumber Industry

Prior to the multivariate analysis of site location factors non-response bias was calculated comparing mean responses from the first and the second mailing (Armstrong and Overton 1977). Results from t-tests for equality of means provides evidence of no statistically significant differences at α=0.05 (Table 22).

Table 22. Mean differences, degrees of freedom, standard errors and tests-statistics for equality of means between responses from first and second mailing.

Variable	Mean	df	Std. Error	t	P-value (2-tail)
Cost of land	-0.316	77	0.303	-1.043	0.300
Cost of logs	-0.061	79	0.164	-0.371	0.711
Sufficient supply of logs	-0.222	79	0.139	-1.599	0.114
Cost of energy	-0.014	79	0.21	-0.065	0.948
Sufficient supply of energy	-0.089	77	0.251	-0.355	0.723
Regional average wages	0.047	78	0.202	0.231	0.818
Non-skilled labor availability	-0.307	78	0.234	-1.310	0.194
Skilled labor availability	-0.020	78	0.193	-0.104	0.918
Quality of roads	0.010	79	0.198	0.05	0.961
Rail and railcar availability	0.458	52	0.417	1.099	0.277
Proximity to ports	-0.220	78	0.302	-0.728	0.469
Distance to markets	-0.023	78	0.268	-0.086	0.932
Proximity to log supply area	-0.233	78	0.183	-1.275	0.206
Trucks and trucking availability	-0.078	79	0.176	-0.445	0.658
Lack of competition from other sawmills	-0.335	79	0.251	-1.337	0.185
Proximity to a university for research support	0.037	79	0.247	0.151	0.880
Availability of technical training for workers	-0.151	79	0.277	-0.546	0.587
Availability of State business incentives	0.369	79	0.282	1.307	0.195

(Table 22 continued)

Variable	Mean	df	Std. Error	t	P-value (2-tail)
Favorable environmental regulations	0.245	79	0.231	1.060	0.293
Favorable state property taxes	0.210	79	0.251	0.837	0.405
Favorable local property taxes	0.210	79	0.251	0.837	0.405
Favorable State fuel taxes	0.145	79	0.236	0.616	0.540
Proximity to health care services	0.069	79	0.237	0.289	0.773
Quality of education for worker's families	0.062	78	0.236	0.264	0.793
Years of operation	6.229	79	13.631	0.457	0.649
2005 annual production (MBF)	-5.647	79	5.906	-0.956	0.342
Maximum distance logs are procured (miles)	-5.376	75	31.907	-0.168	0.867
Maximum distance that finished lumber is shipped	-390.820	56	282.899	-1.381	0.173

Summary results for the responses on the 24 variables included in the interdependence analysis of site attributes in the softwood lumber industry are presented in Table 23. The Likert-scale for this component ranges from 1 (“Not important at all”) to 5 (“Very important”). Based on the variable ratings some of the most important variables that influence the current mill location are the sufficient supply of logs, costs of logs, and proximity to a log supply area. This is not surprising as logs are the primary and a necessary input to sawmills. Their sufficient supply, prices and proximity would be the main drivers for the lumber industry. It is also important to note that these variables show the smallest standard deviations of all. In a second category in terms of their importance there are costs of energy, availability of skilled labor and trucks and trucking availability. These variables correspond to other necessary inputs to the lumber manufacturing process and the means to subsequent product distribution. These are the variables that are expected to be statistically significant in the Conjoint and Spatial Econometrics analyses. It is foreseen that the cost of logs ad distance to logs will be the most important attributes in the Conjoint Analysis, followed by electricity costs, wages and access to roads (as captured by the “ROAD_QUALITY” variable. The appropriateness of the econometric model specification will be partly given by how well it reflects the attributes importance as denoted by individual variable ratings.

Table 23. Summary statistics of the factors that influence current mill location.

		Importance						
Attributes	**n**	**Not Important at All**		**Neither Unimportant Nor Important**		**Very important**	**Mean**	**Std. Deviation**
Cost of land	79	16.5	17.7	20.3	32.9	12.7	3.08	1.299
Cost of logs	81	1.2	1.2	3.7	11.1	82.7	4.73	0.707
Sufficient supply of logs	81	1.2	0.0	3.7	4.9	90.1	4.83	0.608
Cost of energy	81	2.5	2.5	17.3	45.7	32.1	4.02	0.908
Sufficient supply of energy	79	3.8	6.3	20.3	34.2	35.4	3.91	1.076
Regional average wages	80	2.5	6.3	26.3	52.5	12.5	3.66	0.871
Non-skilled labor availability	80	3.8	5.0	27.5	37.5	26.3	3.78	1.018
Skilled labor availability	80	1.3	0.0	16.3	33.8	48.8	4.29	0.830
Quality of roads	81	3.7	3.7	39.5	44.4	8.6	3.51	0.853
Rail and railcar availability	54	22.2	0.0	14.8	27.8	35.2	3.54	1.526
Proximity to ports	80	37.5	16.3	26.3	12.5	7.5	2.36	1.305
Distance to markets	80	10.0	2.5	18.8	47.5	21.3	3.67	1.145
Proximity to log supply area	80	2.5	0.0	3.8	33.8	60.0	4.49	0.795
Trucks and trucking availability	81	0.0	0.0	18.5	34.6	46.9	4.28	0.762
Lack of competition from other sawmills	81	4.9	9.9	32.1	30.9	22.2	3.56	1.095
Proximity to a university for research support	81	37.0	25.9	25.9	9.9	1.2	2.12	1.065
Availability of technical training for workers	81	17.3	16.0	28.4	32.1	6.2	2.94	1.197
Availability of State business incentives	81	17.3	7.4	34.6	29.6	11.1	3.10	1.231
Favorable environmental regulations	81	2.5	4.9	21.0	34.6	37.0	3.99	1.006
Favorable state property taxes	81	6.2	3.7	24.7	38.3	27.2	3.77	1.087
Favorable local property taxes	81	6.2	3.7	24.7	38.3	27.2	3.77	1.087
Favorable State fuel taxes	81	3.7	6.2	30.9	35.8	23.5	3.69	1.020
Proximity to health care services	81	6.2	6.2	29.6	43.2	14.8	3.54	1.025
Quality of education for worker's families	80	7.5	8.8	32.5	43.8	7.5	3.35	1.008

It is relevant to point out the small number of respondents that rated the importance of rail and railcar availability to current sawmill location. Because of the small number of observations this variable is dropped from the model and any further analyses.

Before common factor analysis was carried out a correlation matrix was estimated to determine at first sight the appropriateness of the factor analysis. It was identified that the variables representing favorable state and local property taxes are perfectly correlated, and hence, one was dropped from the analysis (local property taxes). Also, the level of linear correlation between the importance of favorable fuel taxes and state/local property taxes is particularly high at 0.83 and for this reason was also dropped from the model. It is important to note that the process of reducing the number of variables in the model is necessary because of a relatively small sample size (n=81). Common factor analysis was carried out with the remaining 21 variables.

Inspection of the correlation matrix reveals that 141 of the 210 correlations (67 percent) are significant at the 0.05 level of statistical significance. A Barlett's test for Sphericity was ran to determine the overall significance of the correlation matrix. The test is statistically significant with p-value less than 0.001 suggesting that, when taken overall, correlations differ from 0 (no correlation). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (0.766) falls in the acceptable range considered by Hair et al. (1998) of 0.50 or higher (Table 24).

Table 24. Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test for the factor analysis for sawmills site location

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.766
Bartlett's Test of Sphericity	Approx. Chi-Square	671.801
	df	210
	Sig.	<0.001

The criterion used for the number of factors to extract is the latent root criterion. The logic behind this criterion is that any individual factor should account for the variance of at least

a single variable in order to be considered important to explain variance in the model and to be retained for interpretation (Hair et al., 1998). Only factors having eigenvalues greater than 1 are considered significant in the model while others are dismissed.

Of the total 21 variables included in the model a total of 6 common factors were selected based on the eigenvalue criterion. These 6 factors explain a total 65.449 percent of the variance present in the dataset (Table 25).

Table 25. Total variance explained by Common Factors Components

Component	Total	Percent of Variance	Cumulative Percent
1	6.420	30.570	30.570
2	2.032	9.678	40.249
3	1.574	7.496	47.744
4	1.405	6.689	54.433
5	1.173	5.587	60.020
6	1.140	5.429	65.449
7	0.946	4.504	69.953
8	0.885	4.216	74.169
9	0.827	3.938	78.107
10	0.785	3.740	81.847
11	0.681	3.241	85.088
12	0.568	2.705	87.793
13	0.465	2.212	90.005
14	0.407	1.937	91.942
15	0.374	1.783	93.725
16	0.310	1.476	95.201
17	0.280	1.335	96.536
18	0.207	0.987	97.523
19	0.197	0.936	98.459
20	0.181	0.862	99.321
21	0.143	0.679	100.000

Extraction Method: Principal Component Analysis.

The factor loadings from the unrotated solution using a Principal Component extraction method are presented in Table 26. There is considerable variation in the loadings of the variables in the 6 selected common factors. The unrotated factor loading matrix does not provide an easy way to interpret the correlation between the variable and the common factor. A factor orthogonal rotation was then done to simplify the factor structure.

Table 27 shows the coefficients for the rotated factor loadings. Hair et al. (1998) suggest factor loadings of at least 0.65 when the sample size is 70, and 0.60 for samples with 85 observations be used to identify those statistically significant. Based on this criterion for selection the variables proximity to a university for research support, availability of technical training for workers and State business incentives, favorable environmental regulations, favorable state property taxes, proximity to health care services and quality of education for worker's families comprise the first common component.

Table 26. Unrotated component score coefficient matrix

	Component					
Variable	1	2	3	4	5	6
Cost of land	0.156	-0.163	0.328	0.294	0.339	-0.361
Cost of logs	0.559	0.331	-0.086	0.481	-0.176	-0.134
Sufficient supply of logs	0.536	0.347	0.094	0.582	-0.242	-0.010
Cost of energy	0.566	0.220	-0.535	-0.006	-0.055	0.195
Sufficient supply of energy	0.525	0.293	-0.401	-0.006	-0.098	0.497
Regional average wages	0.556	0.297	-0.216	-0.434	0.061	-0.290
Non-skilled labor availability	0.508	0.121	0.237	-0.365	-0.319	-0.370
Skilled labor availability	0.642	0.150	0.396	-0.433	-0.109	0.039
Quality of roads	0.316	-0.331	0.332	-0.031	-0.061	0.350
Proximity to ports	0.306	-0.062	0.480	0.289	-0.197	0.216
Distance to markets	0.575	0.035	0.364	0.031	-0.267	0.092
Proximity to log supply area	0.253	0.669	0.274	0.013	0.345	0.039
Trucks and trucking availability	0.670	0.077	0.173	-0.283	0.039	0.169
Lack of competition from other sawmills	0.528	0.519	-0.020	0.005	0.387	-0.053
Proximity to a university for research support	0.520	-0.219	0.050	0.099	0.584	0.023
Availability of technical training for workers	0.688	-0.396	-0.003	-0.165	0.041	0.123
Availability of State business incentives	0.598	-0.428	-0.114	0.004	0.030	0.069
Favorable environmental regulations	0.702	-0.173	-0.214	0.035	-0.176	-0.314
Favorable state property taxes	0.616	-0.333	-0.314	0.125	-0.109	-0.276
Proximity to health care services	0.753	-0.265	0.008	0.128	0.085	-0.159
Quality of education for worker's families	0.603	-0.283	-0.080	0.055	0.247	0.229

The second common factor includes regional average wages, non-skilled labor availability and skilled labor availability. Cost of logs and sufficient supply of logs comprise the third component. Proximity to log supply area and lack of competition from other sawmills make the fourth component. Proximity to ports makes the fifth component, but given the small sample size the factor loadings for quality of roads and proximity to markets may be considered

significant. The final factor is comprised of costs of land, cost of energy and sufficient supply of energy.

Table 27. Varimax rotated component coefficient matrix

	Component					
Variable	1	2	3	4	5	6
Cost of land	0.251	-0.039	0.154	0.190	0.041	-0.603*
Cost of logs	0.172	0.119	0.782*	0.202	0.038	0.106
Sufficient supply of logs	0.080	0.052	0.832*	0.207	0.254	0.075
Cost of energy	0.374	0.126	0.305	0.179	-0.147	0.627*
Sufficient supply of energy	0.248	0.028	0.245	0.244	0.108	0.765*
Regional average wages	0.263	0.641*	0.060	0.326	-0.283	0.224
Non-skilled labor availability	0.088	0.809*	0.153	0.020	0.110	-0.055
Skilled labor availability	0.181	0.689*	-0.026	0.298	0.430	0.102
Quality of roads	0.292	0.035	-0.103	-0.056	0.588*	0.024
Proximity to ports	0.059	0.017	0.288	0.023	0.631*	-0.110
Distance to markets	0.180	0.376	0.299	0.084	0.523*	0.043
Proximity to log supply area	-0.178	0.136	0.167	0.789*	0.085	0.003
Trucks and trucking availability	0.359	0.439	0.004	0.327	0.344	0.219
Lack of competition from other sawmills	0.191	0.207	0.249	0.733*	-0.091	0.118
Proximity to a university for research support	0.674*	-0.079	-0.028	0.419	0.090	-0.166
Availability of technical training for workers	0.704*	0.274	-0.034	-0.008	0.290	0.133
Availability of State business incentives	0.697*	0.128	0.072	-0.094	0.181	0.099
Favorable environmental regulations	0.582*	0.428	0.397	-0.107	-0.074	0.059
Favorable state property taxes	0.671*	0.244	0.351	-0.205	-0.126	0.038
Proximity to health care services	0.703*	0.256	0.300	0.086	0.147	-0.086
Quality of education for worker's families	0.681*	-0.020	0.036	0.170	0.225	0.146

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 12 iterations. *Variables with significant factor loading.

Based on the results from the factor analysis six common factors are proposed to determine the current location of sawmills. Factor analysis groups the variables in factors and then it is left to the researcher to name them. Based on the association of the different attributes I have named them "Policies, regulations & knowledge", "Human Resources", "Primary Resource Input", "Competition", "Accessibility", "Energy and other costs". Figure 9 summarizes the association of each variable with the corresponding common factor. These factors will be used to guide the building of a spatial econometric model for the occurrence of the Softwood Lumber Industry in the next chapter.

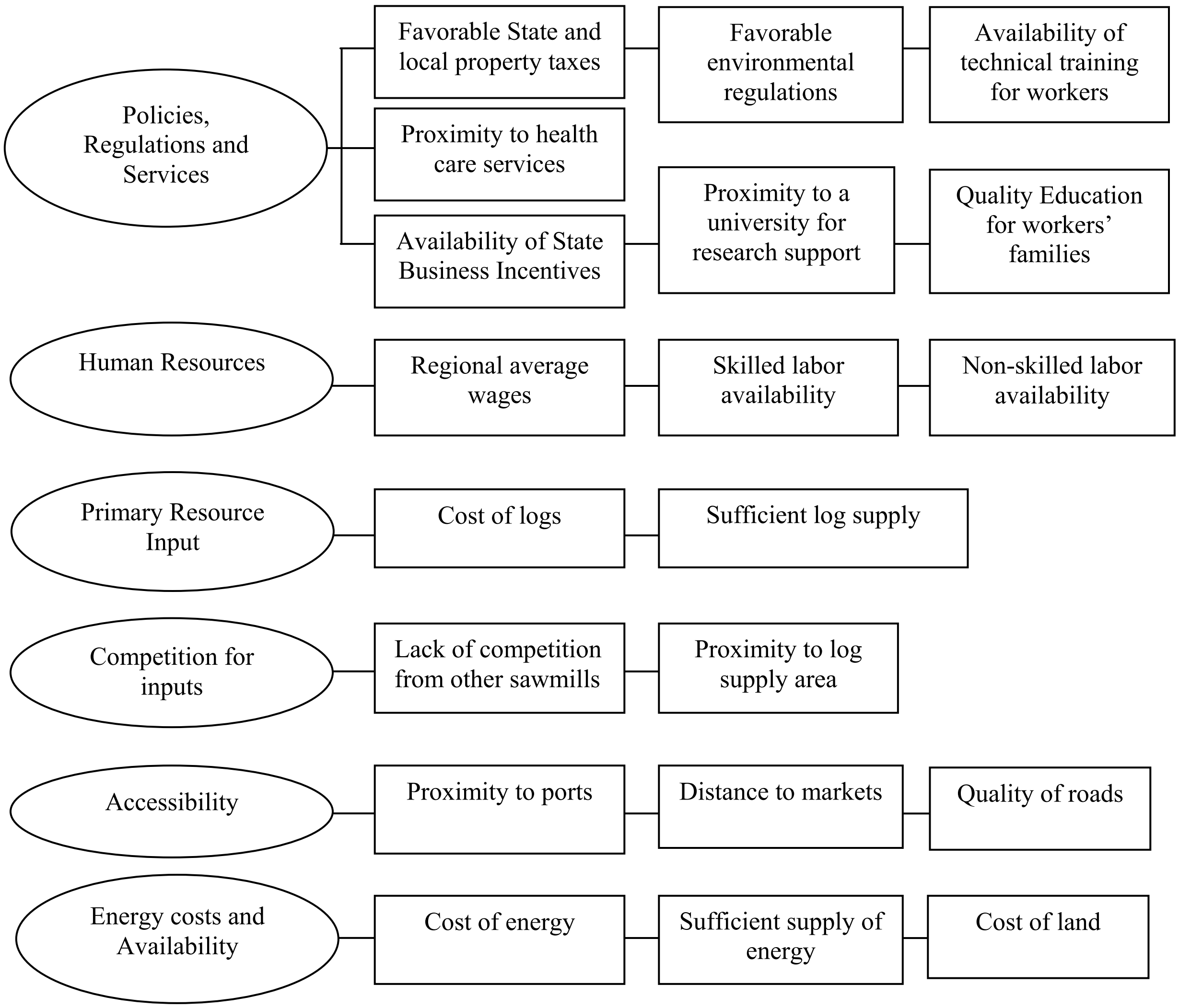


Figure 9. Factors influencing sawmill location based on principal component common factor analysis

6.2 Results of the Conjoint Analysis for the Preferences for Site Attributes for a New Sawmill

Three different model specifications were used to estimate the preferences over the site attributes included in the study. The various models were used because of their different assumptions such as ordinality of preferences in Ordered Logit and Ordered Probit models, and cardinality captured by the Two-limit Tobit (TLT) model. Stata version 9 was used for model estimation and in all analyses presented in this section.

Data was analyzed for the presence of heteroskedasticity using the White and Breusch-Pagan Lagrange Multiplier tests. Both tests failed to reject the null hypothesis of homoskedasticity. The White test-statistic was 1.804, with a Chi-square (7) p-value equal to 0.9699. The Breusch-Pagan test statistic was 1.742 with a Chi-square (7) p-value of 0.9727.

Eight different ratings were provided by every individual respondent which raises the issue of a cluster effect. This cluster specific effect is captured in the error term as it is discussed by Wooldridge (2003) and a robust covariance matrix specification may be necessary. Thus, cluster robust standard errors are calculated to account for the lack of independence of observations per respondent but allowing for independence, and heterogeneity, across participants. Robust standard errors are adjusted for the 81 surveys returned. A cluster robust standard error matrix is used to allow for potential autocorrelation between responses in the same cluster (respondent in this case) and for non-homogenous variances between clusters (Wooldridge 2003).

Table 28 shows the estimated coefficients as well as standard errors, robust standard errors for clustered data and their correspondent test-statistics and p-values for the Ordered Logit model specification. A total of 644 observations were used in the analysis. The coefficient signs are as expected all having negative values except for the variable representing road quality. There is evidence of statistical significance of all variables at $\alpha=0.05$ except for the DISTANCE_MKT variable. This variable, however, becomes significant when using robust standard errors. A smaller robust standard error can be a source of concern to the model as it may be associated to negative correlation within the cluster or a result of model misspecification[1].

---

[1] Sribney (2005) Stata Corp.

The misspecification can be corrected by including suitable within-cluster predictors as will be done in extended models presented later in this section.

Table 28. Ordered Logit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors.

	Coefficient	Std. error	t	P>|t|		Cluster Robust Standard Error*	t	P>|t|
WAGES	-0.32207	0.07199	-4.47	<0.0001		0.07085	-4.55	<0.0001
LOGS_COST	-1.28739	0.08454	-15.23	<0.001		0.12098	-10.64	<0.001
ELECTRICITY	-0.33044	0.07171	-4.61	<0.001		0.05678	-5.82	<0.001
LAND COST	-0.32799	0.07169	-4.58	<0.001		0.04887	-6.71	<0.001
ROAD_QUALITY	0.22705	0.07210	3.15	0.002		0.05735	3.96	<0.001
DISTANCE_LOGS	-0.47012	0.07237	-6.50	<0.001		0.06218	-7.56	<0.001
DISTANCE_MKT	-0.12555	0.07185	-1.75	0.081		0.05118	-2.45	0.014
n	644							
LR Chi^2 (7)	338.02					**Walds Chi^2**	264.88	
Prob > Chi^2	<0.0001					**Prob > Chi^2**	<0.0001	
Pseudo R^2	0.1379							
Log-likelihood	-1056.52							
Cut thresholds	-3.284736	0.172465				-3.28474	0.184187	
	-1.952536	0.120972				-1.95254	0.174862	
	-0.827292	0.099149				-0.82729	0.151128	
	0.169278	0.093676				0.169278	0.125306	
	1.503507	0.107648				1.503507	0.133475	
	2.834712	0.147629				2.834712	0.177815	

*Robust standard errors adjusted for 81 clusters.

The likelihood ratio chi-square for the standard error model and the Wald's Chi-square statistics provide strong evidence of the statistical significance of the coefficients in the model.

The value of R^2 measures as a goodness-of-fit measure is inappropriate for nonlinear models, and as such reliance should be placed on other indices to validate model results. Results of a Chi-square test should provide firm rejection of the joint null hypothesis that all coefficients of the model are equal to zero (Gensch and Recker 1979) as it is the case for this model specification.

Table 29 shows the estimated coefficients as well as standard errors, robust standard errors for clustered data, and their correspondent test-statistics and p-values for the Ordered Probit model specification. The coefficient signs are as expected all having negative values

except for the variable representing ROAD_QUALITY. There is evidence of statistical significance of all variables at α=0.05 except for the DISTANCE_MKT variable which is marginally significant. As in the Ordered Logit model, this variable shows a smaller standard error when using a cluster robust variance specification. The likelihood ratio chi-square for the standard error model and the Wald's Chi-square statistic provide strong evidence of the statistical significance of the coefficients in this model specification.

Table 29. Ordered Probit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors.

	Coefficient	Std. error	z	P>|z|		Cluster Robust Std. Error*	z	P>|z|
WAGES	-0.19345	0.04153	-4.66	<0.001		0.04044	-4.78	<0.001
LOGS_COST	-0.73051	0.04574	-15.97	<0.001		0.06941	-10.52	<0.001
ELECTRICITY	-0.18715	0.04138	-4.52	<0.001		0.03271	-5.72	<0.001
LAND COST	-0.17562	0.04126	-4.26	<0.001		0.02904	-6.05	<0.001
ROAD_QUALITY	0.13399	0.04141	3.24	0.001		0.03263	4.11	<0.001
DISTANCE_LOGS	-0.26505	0.04159	-6.37	<0.001		0.03575	-7.41	<0.001
DISTANCE_MKT	-0.08006	0.04139	-1.93	0.053		0.02973	-2.69	0.007
n	644							
LR Chi^2 (7)	337.88					**Walds Chi^2**	231.12	
Prob > Chi^2	<0.0001					**Prob > Chi^2**	<0.0001	
Pseudo R^2	0.1379							
Log-likelihood	-1056.52							
Cut thresholds	-1.881172	0.090065				-1.88117	0.101379	
	-1.123336	0.066129				-1.12334	0.099913	
	-0.47685	0.057255				-0.47685	0.086415	
	0.102257	0.055321				0.102257	0.072471	
	0.881604	0.061523				0.881604	0.078200	
	1.642344	0.078616				1.642344	0.102382	

*Robust standard errors adjusted for 81 clusters.

Table 30 shows the estimated coefficients, standard errors, test-statistics and corresponding p-values for the TLT model specification. The coefficient signs are as expected all having negative values except for the variable ROAD_QUALITY. There is evidence of statistical significance of all variables at α=0.05. The likelihood ratio test provides strong evidence of the statistical significance of the coefficients in this model specification.

Table 30. Two-limit Tobit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors.

Variables	Coefficient	Std. Error	z	P>|z|
WAGES	-0.30414	0.06426	-4.73	<0.0001
LOGS_COST	-1.13337	0.06461	-17.54	<0.0001
ELECTRICITY	-0.30342	0.06430	-4.72	<0.0001
LAND COST	-0.27839	0.06431	-4.33	<0.0001
ROAD_QUALITY	0.19924	0.06425	3.10	0.0020
DISTANCE_LOGS	-0.42000	0.06435	-6.53	<0.0001
DISTANCE_MKT	-0.13135	0.06424	-2.04	0.0410
CONSTANT	4.22545	0.06425	65.77	<0.0001
n	644		**Pseudo R^2**	0.1311
LR Chi^2 (7)	335.87		**Log likelihood**	-1113.48
Prob > Chi^2	<0.0001			

Observations summary: 51 left-censored observations at rating<=1; 524 uncensored observations; 69 right-censored observations at rating>=7.

Several measures of goodness-of-fit were calculated after model estimation (Table 31). Log-likelihood Intercept only measures computes the likelihood with all parameters by the intercept constrained to zero. The Log-likelihood Full Model corresponds to the log iteration following convergence. McFadden's R^2, compares a model with just the intercept to a model with all parameters. A value of "0" suggests no difference between the two model specifications and values closer to 1 provide evidence of the deviation of the parameters from 0.

Table 31. Measures of goodness-of-fit for the different models used in the study

	Ordered Logit	Ordered Probit	Two-limit Tobit
Log-Likelihood Intercept Only:	-1225.46	-1225.46	-1281.42
Log-Likelihood Full Model:	-1056.449	-1056.521	-1113.48
McFadden's R^2:	0.138	0.138	0.131
AIC:	3.321	3.321	3.486
BIC:	-1968.22	-1968.076	-1880.02

Akaike's Information Criterion (AIC), for comparison of nested models, suggests a better fit for lower values of the criterion (Long and Freese 2006). The Bayesian Information Criterion (BIC) also can be used as a tool to compare models (both nested and non-nested). Models with higher values of BIC are preferred (Long and Freese 2006). These measures of fitness do not suggest preference for any particular model specification. This is not surprising, and almost

expected, as Harrison et al. (2005) determined when comparing similar models specifications for the analysis of agricultural products (crawfish nuggets, and ostrich meat).

Table 32 shows the relative importance of the attributes included in the analysis for the three model specifications. The attributes relative importance was estimated following Halbrendt et al. (1991), described in the Methods section. Relative importance remains homogenous across the different models. The most important variable is LOG_COST that accounts for more than 40 percent of the importance in the model, assuming the model captures 100 percent of the variables considered when location a new sawmill. DISTANCE_LOGS comes second accounting for 15 percent of the importance in the model. ELECTRICITY, LAND COST and WAGES all account for almost similar shares (~10 percent) of the importance of the attributes in the model. ROAD_QUALITY comes sixth in order of importance representing about 7 percent of importance. DISTANCE_MKT is the attribute with the least relative importance with slightly more than a 4 percent share.

Table 32. Attributes relative importance (%) derived from three different model specifications using a rating scale, reduced model.

	Models		
Attributes	**Ordered Logit**	**Ordered Probit**	**Two-limit Tobit**
WAGES	10.42	10.96	10.98
LOGS_COST	41.65	41.37	40.92
ELECTRICITY	10.69	10.60	10.95
LAND COST	10.61	9.95	10.05
ROAD_QUALITY	7.35	7.59	7.19
DISTANCE_LOGS	15.21	15.01	15.16
DISTANCE_MKT	4.06	4.53	4.74
Total	100.00	100.00	100.00

The previous models using only seven explanatory variables raised a flag because of the smaller standard errors with a robust standard error specification – compared to regular standard errors- which again may suggest a problem in the selection of variables and negative correlation. Other model specifications will be tested to account for differences in preferences by people in different positions and the two regions included in the study.

Harrison and Sambidi (2004) surveyed CEO's in the Broiler Industry in the U.S. under the assumption that these are the decision makers when it comes to locate a new operation. The lumber industry however has a different structure and most importantly, facilities are often family-owned which makes it easier to identify the person who finally makes a business decision. This study received responses from sawmill Owners and people in other positions including Sales Managers, Plant Managers and other positions. A binary variable was created to differentiate between sawmill Owners and other types of respondents (Owner=1, 0=otherwise). Also, interactions with the seven attributes included in the model are generated to determine differences in specific coefficients.

Table 33 shows the estimated coefficients as well as standard errors, robust standard errors for clustered data and their correspondent test-statistics and p-values for the augmented ordered logit model specification. The coefficient signs are as expected all having negative values except for the variable representing ROAD_QUALITY. There is evidence of statistical significance of all variables at $\alpha=0.05$ except for the DISTANCE_MKT variable of the variables in the reduced model. The binary variables to account to a pure "Owner" effect is non-significant and of all the interaction terms only WAGE_O and LOGS_O are significant at $\alpha=0.05$. It is also important to notice the congruency between the p-values obtained with the regular standard error and the cluster robust error specification. The likelihood ratio chi-square for the standard error model and the Wald's Chi-square statistic provide strong evidence of the statistical significance of the coefficients in this model specification.

The same model with OWNER and interaction variables is fit under an Ordered Probit specification. Model estimates in Table 34 show the same results as in the Ordered Logit model. Again is also important to notice the agreement between the p-values obtained with the regular standard error and the cluster robust error specification.

Table 33. Ordered Logit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors for an augmented model with “Owner” binary variable and interactions.

	Coefficient	Std. error	z	P>|z|		Cluster Robust Std. Error*	z	P>|z|
WAGES	-0.23894	0.08373	-2.85	0.0040		0.08491	-2.81	0.0050
LOGS_COST	-1.44659	0.09744	-14.85	<0.001		0.14621	-9.89	<0.001
ELECTRICITY	-0.35325	0.08366	-4.22	<0.001		0.07258	-4.87	<0.001
LAND COST	-0.27101	0.08355	-3.24	0.001		0.05378	-5.04	<0.001
ROAD_QUALITY	0.25874	0.08385	3.09	0.002		0.06318	4.09	<0.001
DISTANCE_LOGS	-0.45074	0.08412	-5.36	<0.001		0.07254	-6.21	<0.001
DISTANCE_MKT	-0.08084	0.08382	-0.96	0.335		0.05838	-1.38	0.166
WAGE_O	-0.33473	0.16037	-2.09	0.037		0.15581	-2.15	0.032
LOGS_O	0.51065	0.16079	3.18	0.001		0.21093	2.42	0.015
ELEC_O	0.06247	0.16014	0.39	0.696		0.10477	0.60	0.551
LAND_O	-0.23669	0.16046	-1.48	0.140		0.11863	-2.00	0.046
Q_ROADS_O	-0.09083	0.16027	-0.57	0.571		0.15367	-0.59	0.554
D_LOGS_O	-0.10971	0.16026	-0.68	0.494		0.14421	-0.76	0.447
D_MKT_O	-0.17180	0.16019	-1.07	0.283		0.11632	-1.48	0.140
OWNER	-0.15286	0.16023	-0.95	0.340		0.23351	-0.65	0.513
n	644							
LR Chi^2 (15)	357.88					**Wald Chi^2 (15)**	293.78	
Prob > Chi^2	<0.0001					**Prob > Chi^2**	<0.0001	
Pseudo R^2	0.1467							
Log-likelihood	-1046.521							
Cut thresholds	-3.373972	0.1794338				-3.37397	0.20688	
	-2.018749	0.129281				-2.01875	0.18840	
	-0.871855	0.1096917				-0.87186	0.15389	
	0.1366929	0.1048483				0.13669	0.13816	
	1.489653	0.1170128				1.48965	0.15106	
	2.846986	0.1542156				2.84699	0.19966	

*Standard errors adjusted for 81 clusters.

Table 35 shows the estimated coefficients as well as standard errors and their correspondent test-statistics and p-values for the augmented Two-limit Tobit specification. The results of this model provide more evidence of the significance of the seven attributes in the model except for variable D_MKT and the interaction terms WAGE_O and LOGS_O. These results are in accordance to what determined in the descriptive statistics presented in Table 23 that determine that distance to markets is the least important of all variables. This finding is also in line with Location theory that considers resource-dependent firms to be placed near resource inputs while farther from markets.

Table 34. Ordered Probit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors for an augmented model with "Owner" binary variable and interactions.

	Coefficient	Std. error	z	P>|z|		Cluster Robust Std. Error*	z	P>|z|
WAGES	-0.14633	0.04820	-3.04	0.002		0.04875	-3.000	0.003
LOGS_COST	-0.83055	0.05312	-15.64	<0.001		0.08543	-9.720	<0.001
ELECTRICITY	-0.20415	0.04820	-4.24	<0.001		0.04231	-4.820	<0.001
LAND COST	-0.14217	0.04804	-2.96	0.003		0.03059	-4.650	<0.001
ROAD_QUALITY	0.14766	0.04817	3.07	0.002		0.03465	4.260	<0.001
DISTANCE_LOGS	-0.25423	0.04830	-5.26	<0.001		0.04131	-6.150	<0.001
DISTANCE_MKT	-0.05946	0.04819	-1.23	0.217		0.03306	-1.800	0.072
WAGE_O	-0.19119	0.09342	-2.05	0.041		0.08668	-2.210	0.027
LOGS_O	0.33369	0.09374	3.56	<0.001		0.12054	2.770	0.006
ELEC_O	0.04976	0.09331	0.53	0.594		0.06008	0.830	0.408
LAND_O	-0.13805	0.09345	-1.48	0.140		0.07098	-1.950	0.052
Q_ROADS_O	-0.03917	0.09335	-0.42	0.675		0.08864	-0.440	0.659
D_LOGS_O	-0.06054	0.09334	-0.65	0.517		0.08233	-0.740	0.462
D_MKT_O	-0.08162	0.09331	-0.87	0.382		0.07160	-1.140	0.254
OWNER	-0.08524	0.09335	-0.91	0.361		0.13585	-0.630	0.530
n	644							
LR Chi^2 (15)	359.5					**Wald Chi^2 (15)**	293.78	
Prob > Chi^2	<0.0001					**Prob > Chi^2**	<0.0001	
Pseudo R^2	0.1467							
Log-likelihood	-1045.71							
Cut thresholds	-1.9412	0.1162				-1.9412	0.1162	
	-1.1656	0.1091				-1.1656	0.1091	
	-0.5036	0.0869				-0.5036	0.0869	
	0.0829	0.0778				0.0829	0.0778	
	0.8739	0.0876				0.8739	0.0876	
	1.6501	0.1154				1.6501	0.1154	

*Standard errors adjusted for 81 clusters.

A Chow-like test for the equality between two sets of coefficients (Chow 1960) was used to determine overall significance between the parameter values of Owner and other positions. Results on Table 36 provide strong evidence that the null hypothesis that the binary variable for overall "Owner" differences and interactions with the original explanatory variables are equal to 0 is rejected.

Table 37 shows the relative importance of the attributes included in the analysis for the three model specifications. Relative importance remains homogenous across the different models but differ by respondent position in the company. Still, the most important variable is

LOG_COST that accounts for more than 40 percent of the importance in all models, but it is considered a factor of higher importance to those who do not own a sawmill.

Table 35. Part-worth coefficients of site attributes for a new sawmill under a Two-limit Tobit model specification using a rating scale with "owner" binary variable and interactions.

Variables	Coefficient	Std. Error	z	P>|z|
WAGES	-0.22630	0.07363	-3.07	0.002
LOGS COST	-1.26582	0.07402	-17.10	<0.001
ELECTRICITY	-0.32181	0.07368	-4.37	<0.001
LAND COST	-0.22364	0.07368	-3.04	0.003
ROAD_QUALITY	0.22231	0.07365	3.02	0.003
D_LOGS	-0.39716	0.07372	-5.39	<0.001
D_MKT	-0.09617	0.07362	-1.31	0.192
WAGE_O	-0.29250	0.14333	-2.04	0.042
LOGS_O	0.50893	0.14335	3.55	<0.001
ELEC_O	0.06777	0.14331	0.47	0.636
LAND_O	-0.20638	0.14332	-1.44	0.150
Q_ROADS_O	-0.08112	0.14334	-0.57	0.572
D_LOGS_O	-0.08791	0.14331	-0.61	0.540
D_MKT_O	-0.12842	0.14333	-0.90	0.371
OWNER	-0.14342	0.14334	-1.00	0.317
CONSTANT	4.26257	0.07364	57.88	<0.001
n	644		Pseudo R^2	0.1394
LR Chi^2 (15)	357.37		Log likelihood	-1109.39
Prob > Chi^2	<0.001			

Observations summary: 51 left-censored observations at rating<=1; 524 uncensored observations; 69 right-censored observations at rating>=7.

Table 36. Chow-like Test for the significant differences in responses ratings between company owners and other positions holding remaining explanatory variables constant

Test-statistic	Ordered Logit	Ordered Probit	Two-limit Tobit*
Chi-square (8)	28.00	27.35	2.72
P-value	0.0005	0.0006	0.0059

*For the TLT model, the test-statistic corresponds to an F-test (8, 629).

DISTANCE_LOGS comes second accounting for more than 10 percent of the importance in the model. In the case of owners, however, wages is the second most important characteristics relative to the other attributes. ELECTRICITY, LAND COST and WAGES all account for almost shares between 11 and 7 percent of the importance of the attributes in the model. ROAD_QUALITY comes sixth in order of importance representing about 7 percent of importance. DISTANCE_MKT is the attribute with the least relative importance with slightly

more than a 5 percent share for owners and 3 percent for non-owners. The results are congruent with the descriptive statistics shown in Table 23. These findings also concur with the predictions of Location theory that argues that resource-based industries will place the highest importance on their main resource input (including costs and access to) and they tend to locate closer to those sources rather than to final output markets.

Table 37. Attributes relative importance (%) derived from three different model specifications using a rating scale, reduced model.

	Models					
	Ordered Logit		Ordered Probit		Two-limit Tobit	
Attributes	Owner	Non-owner	Owner	Non-owner	Owner	Non-owner
WAGES	12.4	7.7	12.6	8.2	12.6	8.2
LOGS_COST	42.4	46.7	43.5	46.5	43.0	46.0
ELECTRICITY	9.0	11.4	9.5	11.4	9.4	11.7
LAND COST	11.0	8.7	10.5	8.0	10.4	8.1
ROAD_QUALITY	7.6	8.3	7.0	8.3	7.4	8.1
DISTANCE_LOGS	12.1	14.5	11.8	14.2	11.8	14.4
DISTANCE_MKT	5.5	2.6	5.3	3.3	5.4	3.5
Total	100.0	100.0	100.0	100.0	100.0	100.0

Another research question in addition to the significance of the selected attributes on the preference for the location for a new sawmill is whether there are regional differences. To address this question, a binary variable called SOUTH was created. If an observation comes from a company located in the U.S. South Region it takes on value equal to 1 and 0 otherwise. Interaction terms were generated for the new binary variable and the original explanatory variables. Results for this new model are presented on Table 38 for the Ordered Logit model, Table 39 for the Ordered Probit model specification and Table 40 for the Two-limit Tobit model. None of the new explanatory variables, regional dummy nor interaction terms, are statistically significant at α=0.05. This means that there are regional differences in terms of preferences for the site of a new sawmill. This can be a result of homogenous manufacturing technologies across firms in the lumber industry regardless of their location. If the cost structure of firms between

these two regions is the same then the relative importance placed on each attribute should not differ from one region to the other.

Table 38. Ordered Logit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors for an augmented model with "Regional" binary variable and interactions.

	Coefficient	Std. error	z	P>|z|		Cluster Robust Std. Error*	z	P>|z|
WAGES	-0.20116	0.09955	-2.02	0.043		0.08426	-2.39	0.017
LOGS_COST	-1.14348	0.10735	-10.65	<0.001		0.13481	-8.48	<0.001
ELECTRICITY	-0.34026	0.09962	-3.42	0.001		0.07167	-4.75	<0.001
LAND COST	-0.38226	0.09994	-3.82	<0.001		0.06944	-5.51	<0.001
ROAD_QUALITY	0.26833	0.10027	2.68	0.007		0.08504	3.16	0.002
DISTANCE_LOGS	-0.48613	0.10031	-4.85	<0.001		0.09168	-5.30	<0.001
DISTANCE_MKT	-0.16503	0.09996	-1.65	0.099		0.08066	-2.05	0.041
WAGE_S	-0.25465	0.14157	-1.8	0.072		0.14569	-1.75	0.080
LOGS_S	-0.31672	0.14171	-2.23	0.025		0.21112	-1.50	0.134
ELEC_S	0.01384	0.14157	0.1	0.922		0.11357	0.12	0.903
LAND_S	0.10874	0.14136	0.77	0.442		0.09647	1.13	0.260
Q_ROADS_S	-0.07548	0.14134	-0.53	0.593		0.11340	-0.67	0.506
D_LOGS_S	0.030205	0.14135	0.21	0.831		0.12263	0.25	0.805
D_MKT_S	0.081771	0.14141	0.58	0.563		0.10023	0.82	0.415
SOUTH	-0.05298	0.14152	-0.37	0.708		0.20440	-0.26	0.795
n	644							
LR Chi^2 (15)	347.63					**Wald Chi^2**	282.9	
Prob > Chi^2	<0.0001					**Prob > Chi^2**	<0.0001	
Pseudo R^2	0.1418							
Log-likelihood	-1051.645							
Cut thresholds	-3.341063	0.1891264				-3.34106	0.217051	
	-1.997268	0.1411521				-1.99727	0.22079	
	-0.8583297	0.1205068				-0.85833	0.198647	
	0.1478219	0.1152277				0.147822	0.172599	
	1.494475	0.12842				1.494475	0.183362	
	2.83445	0.1649326				2.83445	0.215101	

*Standard errors adjusted for 81 clusters.

As it was the case for the Ordered Logit, the Ordered Probit model specification that includes coefficients to allow for regional differences fails to reject the individual null hypothesis that coefficients are different from 0 (Table 39). This provides further evidence that the decision-making process is similar between the two regions and their cost structures are alike. Thus, the importance placed on location attributes does not vary from one region to another.

Table 39. Ordered Probit part-worth estimates of site attributes for a new sawmill including standard and robust standard errors for an augmented model with "Regional" binary variable and interactions.

	Coefficient	Std. error	z	P>|z|		Cluster Robust Std. Error*	z	P>|z|
WAGES	-0.12976	0.05772	-2.25	0.025		0.04696	-2.76	0.006
LOGS_COST	-0.65064	0.06031	-10.79	0.000		0.07406	-8.78	0.000
ELECTRICITY	-0.18858	0.05770	-3.27	0.001		0.04022	-4.69	0.000
LAND COST	-0.20964	0.05775	-3.63	0.000		0.04116	-5.09	0.000
ROAD_QUALITY	0.15680	0.05785	2.71	0.007		0.04883	3.21	0.001
DISTANCE_LOGS	-0.28081	0.05799	-4.84	<0.001		0.05070	-5.54	<0.001
DISTANCE_MKT	-0.08900	0.05777	-1.54	0.123		0.04652	-1.91	0.056
WAGE_S	-0.13380	0.08222	-1.63	0.104		0.08240	-1.62	0.104
LOGS_S	-0.17282	0.08228	-2.10	0.036		0.12251	-1.41	0.158
ELEC_S	0.00041	0.08224	0.00	0.996		0.06485	0.01	0.995
LAND_S	0.06725	0.08217	0.82	0.413		0.05619	1.20	0.231
Q_ROADS_S	-0.04499	0.08216	-0.55	0.584		0.06509	-0.69	0.489
D_LOGS_S	0.02897	0.08216	0.35	0.724		0.06824	0.42	0.671
D_MKT_S	0.01664	0.08215	0.20	0.839		0.05895	0.28	0.778
SOUTH	-0.02779	0.08220	-0.34	0.735		0.11814	-0.24	0.814
n	644							
LR Chi^2 (15)	346.15					**Wald Chi^2**	236.33	
Prob > Chi^2	<0.0001					**Prob > Chi^2**	<0.0001	
Pseudo R^2	0.1412							
Log-likelihood	-1052.3869							
Cut thresholds	-1.90980	0.10010				-1.9098	0.1215	
	-1.14598	0.07801				-1.1460	0.1271	
	-0.49313	0.06989				-0.4931	0.1151	
	0.09120	0.06795				0.0912	0.1005	
	0.87545	0.07338				0.8755	0.1048	
	1.63860	0.08897				1.6386	0.1229	

*Standard errors adjusted for 81 clusters.

The Two-limit Tobit model specification (Table 40) including coefficients to allow for regional differences also fails to reject the individual null hypothesis that coefficients are different from 0. Similar results from all three model specifications provide strong evidence in favor of homogenous preferences between regions.

A Chow-like test for the equality between two sets of coefficients (Chow 1960) was performed. Results on Table 41 provide strong evidence that the null hypothesis that the binary variable for overall regional differences and interactions with the original explanatory variables are equal to 0 fails to be rejected.

Table 40. Part-worth coefficients of site attributes for a new sawmill under a Two-limit Tobit model specification including terms for regional differences and interactions.

Variables	Coefficient	Std. Error	z	P>|z|
WAGES	-0.20478	0.08934	-2.290	0.022
LOGS_COST	-0.99881	0.08953	-11.160	0.000
ELECTRICITY	-0.29925	0.08938	-3.350	0.001
LAND COST	-0.32817	0.08940	-3.670	0.000
ROAD_QUALITY	0.23002	0.08934	2.570	0.010
D_LOGS	-0.44021	0.08941	-4.920	0.000
D_MKT	-0.13968	0.08934	-1.560	0.118
WAGE_S	-0.20378	0.12764	-1.600	0.111
LOGS_S	-0.27275	0.12765	-2.140	0.033
ELEC_S	-0.00772	0.12763	-0.060	0.952
LAND_S	0.10199	0.12763	0.800	0.425
Q_ROADS_S	-0.06300	0.12763	-0.490	0.622
D_LOGS_S	0.04265	0.12763	0.330	0.738
D_MKT_S	0.01716	0.12763	0.130	0.893
SOUTH	-0.04686	0.12763	-0.370	0.714
CONSTANT	4.24818	0.08934	47.550	0.000
n	644			
LR Chi^2 (15)	344.06			
Prob > Chi^2	<0.0001			
Pseudo R^2	0.13430			
Log likelihood	-1109.39			

Observations summary: 51 left-censored observations at rating<=1; 524 uncensored observations; 69 right-censored observations at rating>=7.

Table 41. Chow-like test for the significant difference in model coefficients between the U.S. South and the Pacific Northwest regions.

Test-statistic	Ordered Logit	Ordered Probit	Two-limit Tobit*
Chi-square (8)	9.57	8.26	1.03
P-value	0.2963	0.4083	0.4123

*For the TLT model, the test-statistic corresponds to an F-test (8, 629).

Table 42 shows AIC and BIC criterion for the different model specifications. Based on these measures there is little evidence that the new models with interactions are a better fit than the original 7-attribute model. Absolute differences between AIC values range from 0.021 to 0.06 which corresponds to what Raftery (1996) considers to be weak favoring one model against another. Nevertheless, based on the results of the Chow-like tests the model that allows for Owner differences is identified as the preferred specification and it will be used in further analysis. Furthermore, given the similarity in the behavior of the estimation models, only output

of the Ordered Probit Model is presented. Results for the Tobit Model are presented as well because of its different theoretical basis (interval rating scale versus the ordered preferences).

Table 42. AIC and BIC measure for goodness-of-fit for the different models and specifications.

	AIC	BIC
Simplified Model (no interactions)		
Ordered Logit	3.321	-1968.22
Ordered Probit	3.321	-1968.08
Two-limit Tobit	3.486	-1880.02
Augmented Model with Owner Effects		
Ordered Logit	3.315	-1936.33
Ordered Probit	3.313	-1937.96
Two-limit Tobit	3.477	-1849.78
Augmented Model for Regional Effects		
Ordered Logit	3.331	-1926.09
Ordered Probit	3.333	-1924.6
Two-limit Tobit	3.498	-1836.47

Predicted probabilities for the augmented model including Owner-effects were estimated for the Ordered Probit Model and are presented in Figures 10 and 11. The eight profiles included in the survey are presented along with what I have called an "Ideal" profile that corresponds to a site with the lowest costs, shortest distances to markets and source of logs, and better quality of roads. As expected this "Ideal" profile has the highest probability for the 7th category corresponding to "Very attractive" in the 7-point rating scale.

Figure 10 shows the estimated probabilities for Owners and Figure 11 for Non-owners and are presented in ascending order based on the probability of the profile to fall the highest category. Profiles 5, 6, 7 and 8 have the highest predicted probability of being selected as most attractive by owners. Notice that these four profiles correspond to those for which the average price per log is the lowest level.

Owners' predicted preferences differ from that of those in other positions (non-owner). Profile 2 is associated with the highest probability of falling in the most attractive category. This profile, although having a "high" price for logs has the most preferable conditions for the

remaining variables, except for a higher distance to markets which as presented in the model has no statistical significance. Compared to owners it can be noticed for example how profile 8 is the highest preferred by Owners but only comes third in the predicted probability of falling in the "Very attractive" category. This is due to the higher importance that Owners place in wages. Profile 8 has a lower average hourly wage in the region.

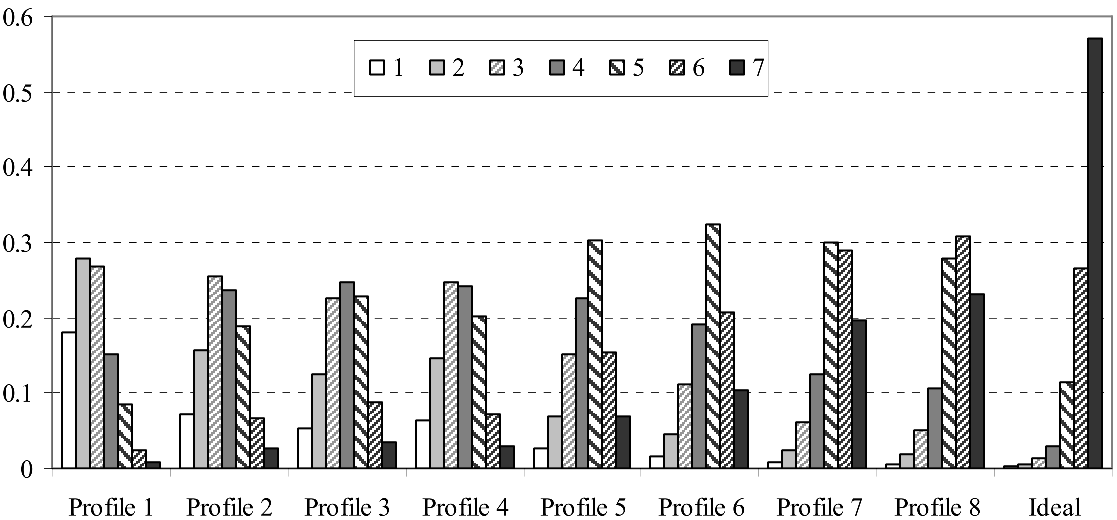


Figure 10. Predicted probabilities for eight different location profiles, and an "Ideal" profile, for "Owners" using an Ordered Probit model.

Predicted profile ratings were estimated for the Two-limit Tobit model and are presented in Figures 12 and 13. Both charts present the estimated rating and lower and higher limits corresponding to a 95 percent confidence interval. It is important to notice the difference in the nature of the model compared to the ordered preferences. Ordered preferences models, such as the Ordered Probit present the probabilities of the outcome falling in each of the possible categories of site attractiveness. However, the Two-limit Tobit model produces a single predicted value that as presented below, includes a confidence interval. The ordering of the profiles does not differ between Owners and Managers for the profiles considered to be more attractive.

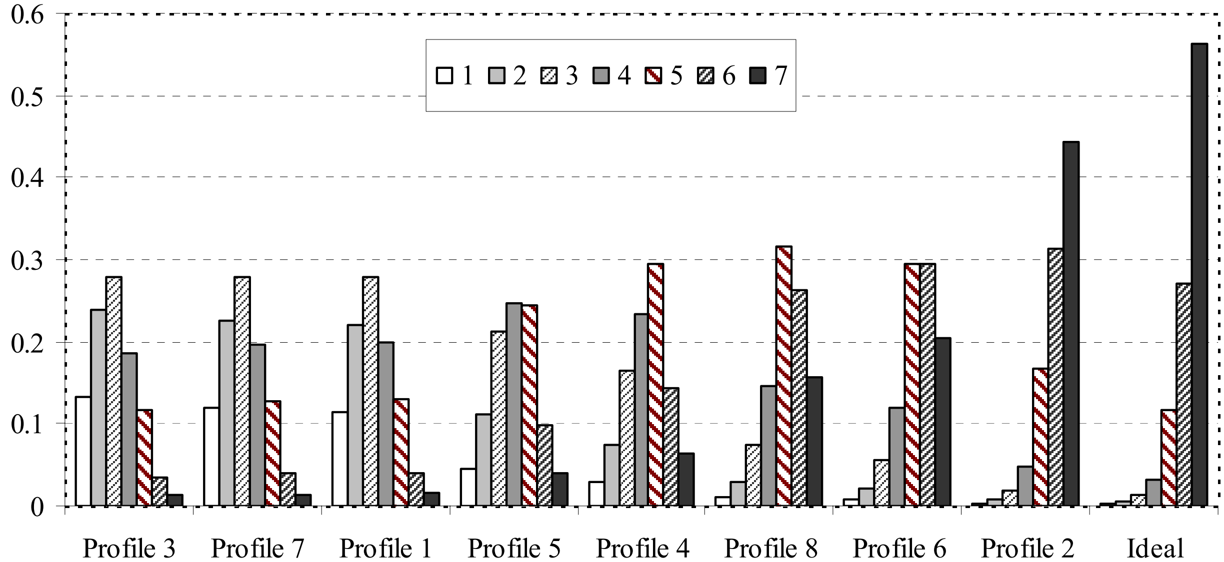


Figure 11. Predicted probabilities for eight different location profiles, and an "Ideal" profile", for "Non-owner" using an Ordered Probit model.

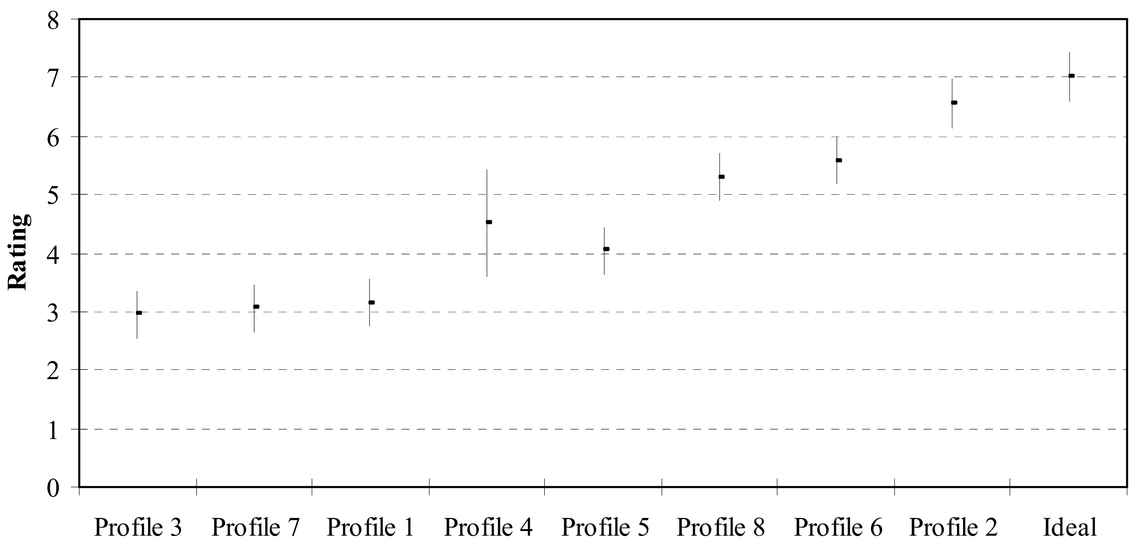


Figure 12. Predicted probabilities for the eight different location profiles, and an "Ideal" profile, for "Owners" using a Two-limit Tobit model.

Non-owners' predicted ratings vary slightly compared to that of the Owners. Also, notice that the non-owners' responses showed a higher degree of variation that result in wider

confidence intervals. Only the preferred order of the three least attractive profiles differs from that of the Owners.

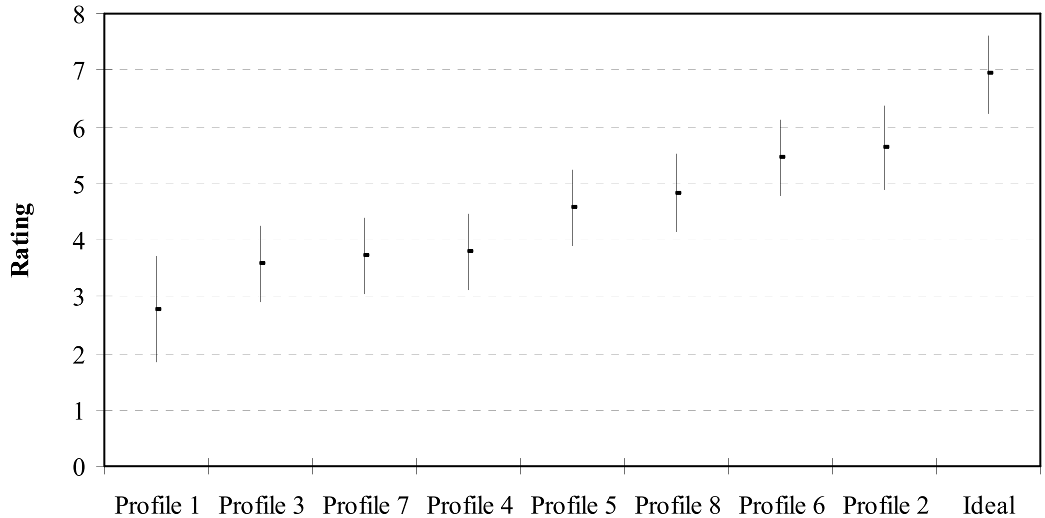


Figure 13. Predicted probabilities for the eight different location profiles, and an "ideal profile", for "Non-owners" using a Two-limit Tobit model.

In non-linear models the analysis of statistical significance of individual coefficient estimates should be accompanied with their corresponding analysis of marginal probability effects to determine their actual effect. Binary models imply that the magnitude of the effect from a particular predictor variable depends on the level of the dependent choice variable and not solely on the estimated coefficients (Gensch and Recker 1979). This is given in the specification of the probability distribution model as changes in the descriptors will have a lesser significant impact on the probability of a binary response when the probability is close to 0 or 1 than when it is around ½ (Theil 1970). This situation deems to be a better representation of real-life situations than a linear regression approach which ignores the level of the variables. Given the proposed models, marginal effects have to be estimated assuming values for the remaining explanatory variables in the model.

First, the marginal effect of the coefficient corresponding to the average price for logs was tested for Owners and Others. This variable was selected because it showed the highest coefficient values of the model and exemplifies the analysis of marginal effects. The average price for logs has a statistically significant effect at all different threshold levels. As expected, this attribute has a lesser significant marginal effect at lower threshold levels. This variable is the most important for a site to be considered very attractive when the other variables in the model are set at levels considered to be appealing as well. Its marginal effect is most significant at higher threshold levels suggesting this variable is crucial to determine a site to be very attractive to decision makers (Table 43).

Table 43. Marginal effects of the average price for logs (LOGS_COST) for a profile with remaining attributes at attractive levels.

Owners	**Coefficient**	**Std. Err.**	**z**	**P>z**
Threshold μ_1	0.02116	0.01344	1.57	0.115
Threshold μ_2	0.10339	0.04499	2.30	0.022
Threshold μ_3	0.26524	0.07724	3.43	0.001
Threshold μ_4	0.45044	0.07935	5.68	<0.001
Threshold μ_5	0.59333	0.03132	18.94	<0.001
Threshold μ_6	0.48001	0.06674	7.19	<0.001
Non-owners				
Threshold μ_1	0.01938	0.00749	2.59	0.010
Threshold μ_2	0.09710	0.02428	4.00	<0.001
Threshold $\mu 3$	0.25426	0.04233	6.01	<0.001
Threshold $\mu 4$	0.43942	0.04613	9.53	<0.001
Threshold $\mu 5$	0.59227	0.03039	19.49	<0.001
Threshold $\mu 6$	0.49005	0.04473	10.96	<0.001

The marginal effect of the average price for logs was estimated when the other site attributes are set at least attractive levels (Table 44). The variable is still very significant at all levels but it has a lesser important effect at higher levels of site attractiveness. This is another sign of the appropriateness of this model to capture site preferences and its ability to model marginal effects.

Table 44. Marginal effects of the average price for logs (LOGS_COST) for a profile with remaining attributes at least attractive levels.

Owners	Coefficient	Std. Err.	z	P>z
Threshold μ_1	0.30720	0.08600	3.57	<0.001
Threshold μ_2	0.51819	0.06217	8.34	<0.001
Threshold μ_3	0.54742	0.05023	10.90	<0.001
Threshold μ_4	0.42580	0.08394	5.07	<0.001
Threshold μ_5	0.19287	0.07154	2.70	0.007
Threshold μ_6	0.05265	0.02943	1.79	0.074
Non-owners	**Coefficient**	**Std. Err.**	**z**	**P>z**
Threshold μ_1	0.35689	0.05143	6.94	<0.001
Threshold μ_2	0.54280	0.04388	12.37	<0.001
Threshold μ_3	0.52547	0.04551	11.55	<0.001
Threshold μ_4	0.37807	0.04946	7.64	<0.001
Threshold μ_5	0.15375	0.03318	4.63	<0.001
Threshold μ_6	0.03761	0.01223	3.08	0.002

Table 45 shows the marginal probability effects due to a respondent being an Owner, compared to a person in another position. The level of significance varies depending on the combination of log prices and conditions of the remaining attributes in the model. Table 45 shows the results of the marginal analysis under four different scenarios (low price for logs/least attractive conditions, high price for logs/least attractive conditions, low price for logs/most attractive conditions, high price for logs/most attractive conditions) evaluated at three different threshold levels.

Table 45. Marginal effects of the respondent being a company owner considering different levels of the average price for logs (LOGS_COST) for a profile with remaining attributes at most and least attractive levels.

Threshold	Cost of logs	Other attributes	Coefficient	Std. error	z	P>z
Threshold μ_5	\$46.31/ton	Least attractive	-0.12327	0.03783	-3.26	0.001
Threshold μ_5	\$62.25/ton	Least attractive	0.00233	0.00494	0.47	0.637
Threshold μ_5	\$46.31/ton	Most attractive	-0.00977	0.07212	-0.14	0.892
Threshold μ_5	\$62.25/ton	Most attractive	0.22594	0.09799	2.31	0.021
Threshold μ_3	\$46.31/ton	Least attractive	-0.31130	0.09730	-3.20	0.001
Threshold μ_3	\$62.25/ton	Least attractive	0.03132	0.06169	0.51	0.612
Threshold μ_3	\$46.31/ton	Most attractive	-0.00111	0.00828	-0.13	0.894
Threshold μ_3	\$62.25/ton	Most attractive	0.16112	0.06015	2.68	0.007
Threshold μ_1	\$46.31/ton	Least attractive	-0.12018	0.05581	-2.15	0.031
Threshold μ_1	\$62.25/ton	Least attractive	0.06250	0.11569	0.54	0.589
Threshold μ_1	\$46.31/ton	Most attractive	-0.00002	0.00011	-0.13	0.895
Threshold μ_1	\$62.25/ton	Most attractive	0.01596	0.00739	2.16	0.031

There is a clear trend denoting statistically significant differences in the preferences of "Owners" compared to "Non-owners". Table 45 shows that the coefficient for the differences in probabilities is negative which indicates the probability of preferences for "Owners" are higher than "Non-owners" when the cost of logs is lower. Further, this difference is significant when conditions for the remaining attributes are least attractive. There is the opposite case for people with positions other than the owner. For "Non-owner" marginal effects of the costs of logs are statistically significant higher, as indicated by positive coefficients, when the conditions are more attractive.

Table 46 summarizes the differences between owners and non-owners' preferences in regard to their statistical significance. Notice that significant (α=0.05) for both extreme prices but depending on the conditions of the remaining variables. When the remaining attributes are at their most attractive levels, high log prices are significant and more important to non-owners. The inverse is true at the other end. When the remaining attributes are at their least attractive levels, low log prices are significant and more important to company owners.

Table 46. Observed differences and significance between owners and non-owners' preferences at different average price for logs (LOGS_COST) and remaining attributes at most and least attractive levels.

Price per logs/ remaining attributes	Low price per logs ($46.31/ton)	High price per logs ($62.25/ton)
Most attractive site attributes	Non-significant	Significant (more attractive to non-owners)
Least attractive site attributes	Significant (more attractive to Owners)	Non-significant

6.3 Conditional Logit (Fixed Effects) Model for the Analysis of Choice-based Responses

A Conditional Logit, also known as fixed-effects logit model, was used to study the responses from the choice-based section of the questionnaire (Appendix 1). The model developed my McFadden (1974, 1986) breaks down the utilities derived from the location

attributes. Part-worth coefficients are estimated under the assumption that the profile of choice maximizes the respondent's utility.

A first model was fit for the overall respondents without interaction variables and includes a total 642 observations comparing two profiles at the time (Appendix 1). Results are presented in Table 47. A first look at the results shows that the coefficients carry the expected correct signs for the variables WAGES, LOGS COST, ELECTRICITY, ROADS_QUALITY, D_LOGS, and D_MKT. However, the variable LAND COST has a positive sign which is contrary to what would be expected from a relationship between the probabilities of selecting a site with higher land costs. All variables expect for ROADS_QUALITY are statistically significant at α=0.05.

Table 47. Conditional Logit part-worth coefficients for site attributes for the reduced model specification using Choice-based responses.

	Coefficient	Std. error	z	P>|Z|
WAGES	-0.66112	0.13666	-4.84	<0.001
LOGS COST	-1.01706	0.13574	-7.49	<0.001
ELECTRICITY	-0.64342	0.13717	-4.69	<0.001
LAND COST	1.12098	0.13554	8.27	<0.001
ROAD_QUALITY	0.10391	0.13894	0.75	0.455
D_LOGS	-0.45985	0.13731	-3.35	0.001
D_MKT	-0.55721	0.13672	-4.08	<0.001
n	642			
Log likelihood	160.95382			
LR $Chi^2(7)$	363.1			
Prob > Chi^2	<0.001			
Pseudo R^2	0.5301			

Table 48 shows the result for an expanded model that includes interactions for respondents who are sawmill owners. These interactions are included because of their significance determined in the previous section. Under this new model the variable LAND_COST still has the wrong sign, and the variable capturing road quality effects is marginally significant at α=0.05. None of the interactions terms is significant at α=0.05. In

addition to obtaining the variable LAND_COST with the wrong sign the absolute value for the coefficient is the highest among all variables.

Table 48. Conditional Logit part-worth estimates of site attributes for a new sawmill for an augmented model specification with "owner" binary variable and interactions.

	Coefficient	Std. error	z	P>|t|
WAGES	-0.95465	0.21247	-4.49	<0.001
LOGS COST	-0.98237	0.21101	-4.66	<0.001
ELECTRICITY	-0.84531	0.21344	-3.96	<0.001
LAND COST	1.39117	0.21147	6.58	<0.001
ROAD_QUALITY	0.40879	0.21495	1.90	0.057
D_LOGS	-0.43652	0.21289	-2.05	0.040
D_MKT	-0.54585	0.21189	-2.58	0.010
WAGE_O	0.63782	0.32947	1.94	0.053
LOGS_O	-0.14577	0.32535	-0.45	0.654
ELEC_O	0.24818	0.32888	0.75	0.450
LAND_O	-0.54885	0.32617	-1.68	0.092
Q_ROADS_O	-0.69463	0.33258	-2.09	0.037
D_LOGS_O	-0.08896	0.32904	-0.27	0.787
D_MKT_O	-0.05681	0.32915	-0.17	0.863
n	642		**Prob > Chi^2**	<0.001
Log likelihood	155.4695		**Pseudo R^2**	0.5461
LR Chi2(15)	374.07			

These results may suggest that this model specification is not the most appropriate for this problem. The full-profile choice-based approach uses a complete set of factors to describe a product profile and compares two profiles at the same time. The major risk of this method for preference elicitation is the possibility of information overload and as a result, respondents may simplify the experimental task by ignoring variations in the less important factors or by simplifying the factor levels themselves. Green and Srinivasan (1978) warn that full-profiles with more than five or six attributes can result in information overload. Swait and Adamowicz (2001) report that efficiency of responses to choice-based experiment questions can be significantly affected by the length and difficulty of the choice tasks. This may well be what is happening in this problem as the costs of logs and distance to logs account for about 60 percent of all the importance relative to the other five attributes included in the profiles.

The coefficients of the Conditional Logit model are not consistent with logical expectations. Furthermore, the results are not congruent with the initial summary statistics presented in Table 23. It is determined that this model specification is not the most appropriate because of the multivariate nature of the location decision problem could not be captured properly and will not be further discussed.

6.4 Perceptions about Cluster Patterns in the Softwood Lumber Industry

Participants in the study were also asked for their impression that the softwood lumber industry is spatially arranged in cluster patterns. This information is used to provide a stronger case for the use of a spatially correlated observations model in the econometric analysis. The majority of respondents (72.5 percent) consider that softwood sawmills tend to be located in geographical clusters (Table 49).

However, a major proportion of respondents agree that such a spatial clustering is not beneficial to the industry (Table 50). The reasons why respondents consider that such spatial arrangement is not beneficial is due to several causes. A mean response value of 4.3 corresponding to "Strongly Agree" in a 1 to 5 Likert-scale (1=strongly disagree, 5 strongly agree) reveals that respondent strongly agree that industrial clustering contributes to increased log prices and competition (Table 50).

Table 49. Believe in the existence of a cluster pattern in the softwood lumber industry and whether this is beneficial to the industry.

Statement	Yes	No	Do not know
Overall, do you believe that softwood sawmills tend to be located in geographical clusters or groups? (N=80)	58 (72.5%)	10 (12.5%)	12 (15%)
Do you think it is beneficial for softwood sawmills to be located close to each other in a cluster or group? (n=80)	11 (13.8%)	49 (61.3%)	20 (25%)

Since the cost of logs is the most important factor when locating a sawmill, as determined in the previous section, then this factor constitutes a major barrier limiting clustering in the softwood lumber industry. This factor corresponds to one of the centrifugal forces in the Fujita

and Krugman (2004) tradition. On the opposite direction, centripetal forces attract industries. In the case of the softwood lumber industry the respondents identify the availability of more local suppliers and a greater opportunity to vertically integrate into manufacturing secondary products as the two strongest factors favoring industry clustering (Table 50).

Table 50. Summary statistics and proportions regarding the level of agreement with consequences derived from potential softwood lumber mills arrangement in clusters over mills that are spatially dispersed.

		Completely disagree			Completely Agree			
	n	**1**	**2**	**3**	**4**	**5**	**Mean**	**Std. Deviation**
Better access to workers with managerial skills	78	6.4	25.6	32.1	32.1	3.8	3.013	1.000
Larger pool of skilled workers	77	5.2	20.8	23.4	40.3	10.4	3.299	1.077
Larger pool of unskilled labor	78	5.1	28.2	37.2	26.9	2.6	2.936	0.931
Better availability of raw materials	76	25.0	32.9	15.8	14.5	11.8	2.553	1.331
Better able to compete with other regions	78	12.8	38.5	24.4	17.9	6.4	2.667	1.113
Availability of more local suppliers	78	5.1	11.5	17.9	53.8	11.5	3.551	1.015
Potential collaboration among sawmills	75	6.7	22.7	30.7	32.0	8.0	3.120	1.065
Better access to information services	77	6.5	15.6	41.6	32.5	3.9	3.117	0.946
Greater opportunity to vertically integrate into manufacturing secondary products	78	1.3	9.0	32.1	48.7	9.0	3.551	0.832
Greater informal sharing of information between plants	78	7.7	9.0	37.2	38.5	7.7	3.295	1.008
Easier access to investment capital	78	11.5	23.1	50.0	14.1	1.3	2.705	0.899
Improved innovation through increased competition	78	7.7	14.1	20.5	44.9	12.8	3.410	1.122
A better organized industry	77	9.1	15.6	37.7	31.2	6.5	3.104	1.046
Increased energy costs	78	9.0	29.5	50.0	10.3	1.3	2.654	0.835
Increased log prices	79	2.5	5.1	2.5	39.2	50.6	4.304	0.939
Increased labor costs	79	0.0	10.1	12.7	51.9	25.3	3.924	0.888
More congestion on local roads	79	2.5	7.6	25.3	49.4	15.2	3.671	0.916
Increased competition	79	0.0	2.5	5.1	49.4	43.0	4.329	0.693

Completely Disagree (1), Somewhat Disagree (2), Neither Disagree nor Agree (3), Somewhat Agree (4), Strongly Agree (5).

Based on the respondents agreement with the different statements that capture centrifugal and centripetal forces a common factor analysis is performed with the pre-specified hypothesis that these factors can be classified in the above mentioned forces. Before common factor analysis was carried out a correlation matrix was estimated to determine at first sight the appropriateness of the factor analysis. There were no variables that showed a high degree of correlation. Only the ratings to the statements that a cluster pattern results in “Better access to workers with managerial skills” and a “Larger pool of skilled workers” had a relatively high correlation of 0.679, but this value is not considered a major concern.

Inspection of the correlation matrix reveals that 78 of the 153 correlations (51 percent) are significant at the 0.05 level of statistical significance. A Barlett’s test for Sphericity was ran to determine the overall significance of the correlation matrix. The test is statistically significant with p-value less than 0.001 suggesting that, when taken overall, correlations differ from 0 (no correlation). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (0.969) falls in the acceptable range considered by Hair et al. (1998) of 0.50 or higher (Table 51).

Table 51. KMO and Bartlett's test for the appropriateness of factor analysis.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.696
Bartlett's Test of Sphericity	Approx. Chi-Square	498.823
	Degrees of Freedom	153
	Sig.	2.953E-38

Compared to the common factor analysis carried out for the variables affecting sawmill locations, here the criterion used for the number of factors to extract is based on a predetermined number of factors. As discussed in the theoretical framework section there are two major forces believed to affect industry clustering, centrifugal and centripetal forces (Fujita and Krugman 2004). Under this model only two principal components were extracted that explain a total 41.035 percent of the variance in the dataset.

The unrotated component matrix show that the biggest factor loading for the first component corresponds to the centripetal forces of clustering (Table 53). The remaining variables correspond to the centrifugal force working against industry clustering.

An orthogonal rotation was carried out for the original component matrix solution. This was done to facilitate interpretation and redistribute the variance from one factor to the other, achieving a more meaningful pattern (Hair et al. 1998). Table 54 shows the factor loadings for the two components.

Table 52. Total variance explained by Common Factors Components Model to the Softwood Lumber Industry clustering problem.

Component	Total	Percent of Variance	Cumulative Percent
1	4.932	27.401	27.401
2	2.454	13.634	41.035
3	1.589	8.826	49.861
4	1.355	7.529	57.390
5	1.211	6.730	64.120
6	1.010	5.608	69.728
7	0.924	5.133	74.861
8	0.741	4.118	78.979
9	0.634	3.522	82.501
10	0.566	3.144	85.645
11	0.533	2.963	88.608
12	0.468	2.602	91.210
13	0.380	2.114	93.323
14	0.337	1.875	95.198
15	0.310	1.724	96.922
16	0.221	1.230	98.153
17	0.180	0.999	99.152
18	0.153	0.848	100.000

Extraction Method: Principal Component Analysis.

Table 53. Unrotated Factor Component matrix for the Softwood Lumber Industry clustering problem.

By clustering together, softwood lumber mills have the following advantages/disadvantages over mills that are dispersed	1	2
Better access to workers with managerial skills (AX_MANG)	0.658	-0.188
Larger pool of skilled workers (POOL_SKILL)	0.684	-0.130
Larger pool of unskilled labor (POOL_UNSKILL)	0.412	-0.062
Better availability of raw materials (RAW_AVAIL)	0.668	-0.075
Better able to compete with other regions (COMPT_OTHER)	0.769	-0.024
Availability of more local suppliers (LOCAL_SUPP)	0.475	0.254
Potential collaboration among sawmills (COLBOR)	0.542	-0.025

(Table 53 continued)

Better access to information services (INFO_SER)	0.604	0.287
Greater opportunity to vertically integrate into manufacturing secondary products (INTEGRATE)	0.421	0.121
Greater informal sharing of information between plants (SHARE_INFO)	0.461	0.030
Easier access to investment capital (INVEST_K)	0.686	0.271
Improved innovation through increased competition (INNOVAT)	0.555	0.368
A better organized industry (ORGNZD)	0.740	0.263
Increased energy costs (INC_ENERGY)	-0.085	0.477
Increased log prices (INC_LOG)	-0.331	0.678
Increased labor costs (INC_LABOR)	-0.241	0.734
More congestion on local roads (INC_CONGEST)	-0.100	0.604
Increased competition (INC_COMPET)	-0.201	0.600

Extraction Method: Principal Component Analysis.

Table 54. Varimax rotated Factor Component Matrix for Lumber Industry clustering problem.

By clustering together, softwood lumber mills have the following advantages/disadvantages over mills that are dispersed	1	2
Better access to workers with managerial skills (AX_MANG)	0.595	-0.339
Larger pool of skilled workers (POOL_SKILL)	0.634	-0.289
Larger pool of unskilled labor (POOL_UNSKILL)	0.386	-0.158
Better availability of raw materials (RAW_AVAIL)	0.631	-0.232
Better able to compete with other regions (COMPT_OTHER)	0.741	-0.206
Availability of more local suppliers (LOCAL_SUPP)	0.522	0.134
Potential collaboration among sawmills (COLBOR)	0.520	-0.153
Better access to information services (INFO_SER)	0.655	0.136
Greater opportunity to vertically integrate into manufacturing secondary products (INTEGRATE)	0.437	0.018
Greater informal sharing of information between plants (SHARE_INFO)	0.455	-0.080
Easier access to investment capital (INVEST_K)	0.731	0.101
Improved innovation through increased competition (INNOVAT)	0.626	0.226
A better organized industry (ORGNZD)	0.782	0.080
Increased energy costs (INC_ENERGY)	0.030	0.483
Increased log prices (INC_LOG)	-0.161	0.737
Increased labor costs (INC_LABOR)	-0.060	0.770
More congestion on local roads (INC_CONGEST)	0.046	0.611
Increased competition (INC_COMPET)	-0.053	0.630

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 3 iterations.

A plot for the two common factor components depicts a clear representation of the opposite directions of the centrifugal and centripetal forces working against and in favor of industrial clustering (Figure14).

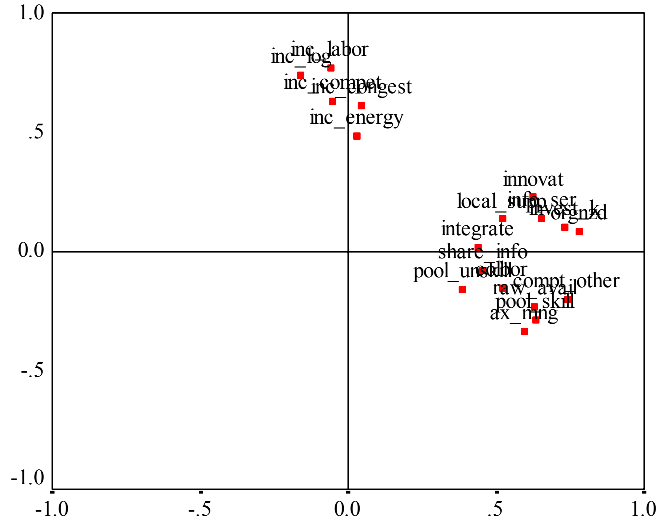


Figure 14. Plot of common factor components for the analysis of centrifugal and centripetal forces in the lumber industry.

The study of preferences for location attributes denotes the importance of the primary input to a resource-based industry as the most important factor. Other variables such as costs of labor, energy, and land are also important but belong to a second category in terms of their importance. Study of preferences reveals that these differ whether a decision maker is an owner or a non-owner. Owners place more importance on the cost of procuring the primary resource input but less weight on wages and energy costs compared to non-owners. There are no regional differences when comparing location preferences for decision-makers in the U.S. Pacific West compared to those in the South. Although industry members consider that the lumber sector clusters in particular geographic areas, there are major forces working against further aggregation. The most important centrifugal forces are increased competition from other sawmills and potentially higher log prices.

CHAPTER 7. RESULTS: EXPLORATORY SPATIAL DATA ANALYSIS

7.1 Description of Data Available on the Location of Sawmills in the U.S. South and Texas

Data for the spatial analysis of the lumber industry comes from the USDA Forest Service, Southern Research Station, SRS-4850 (2005a). A shapefile generated in 2005 describes the types and locations of wood mills in the Southern United States that purchase logs or wood chips for primary processing. The mill location shapefile updates and expands the spatial scope of several earlier datasets generated by the Southern Research Station (2005b). A separate shapefile for the Texas region is also available and is merged with the Southern region data to cover a larger extension of the U.S. territory providing a comprehensive database.

The original datasets include companies manufacturing wood products composites, plywood, veneers, post poles, pulp and lumber. Only sawmills are selected for analysis while companies in other categories were discarded. Exploratory analyses are performed on this data. The Southern and Texas regions are studied separately to determine first-hand evidence of clustering as deviation from complete spatial randomness using inter-point distance tests and a Chi-square for homogenous distribution.

7.2 Deviations from Complete Spatial Randomness in Spatial Distribution of Sawmills in the U.S. South Region

A total of 1,786 sawmills were identified in the U.S. South region by the USDA Forest Service Southern Research Station (2005a). The original dataset includes sawmills in 12 different states (Figure 15). The majority of sawmills concentrate in Tennessee (24 percent), followed by Kentucky (17 percent), Virginia (12 percent), and North Carolina (12 percent). The remaining states host each less than 10 percent of all sawmills identified in the U.S. South. In this classification Oklahoma is considered as another state in the U.S. South.

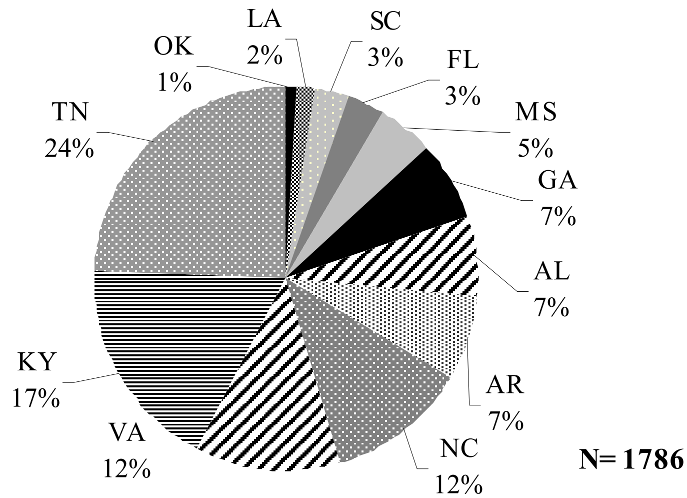


Figure 15. Distribution of sawmills per state in the U.S. South region

Figure 16 presents the spatial distribution of sawmills where each enterprise is identified with a circle. Data is projected in geographic degrees, the longitude coordinates extend from -96.03440 to -75.61914 degrees, and the latitude coordinates extend from 27.21118 to 39.17673 degrees. The distribution of sawmills closely depicts the profile of the U.S. South region.

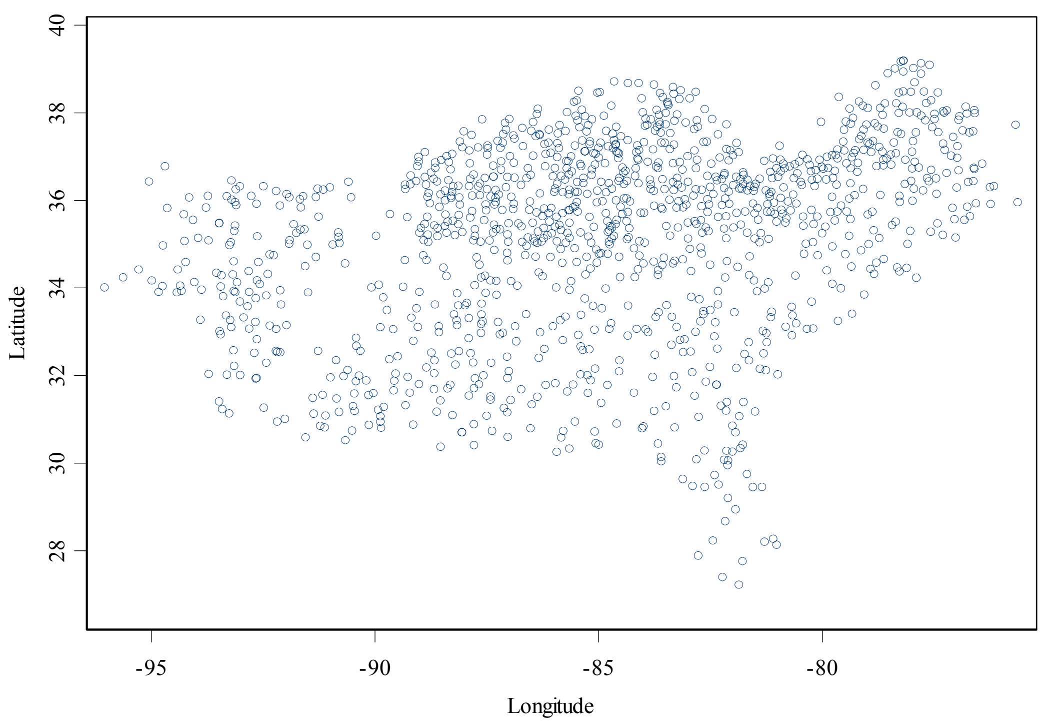


Figure 16. Scatter plot of Sawmills in the U.S. South for the year 2005.

Figure 17 presents the results of the Fhat and Ghat analyses for the study of point data distribution. Both analyses suggest deviation from complete spatial randomness. An excess of long distance neighbors in the Fhat analysis provides evidence of spatial aggregation in the data. On the other hand, a large number of points at short neighbor distances in the Ghat analysis suggest clustering of data points.

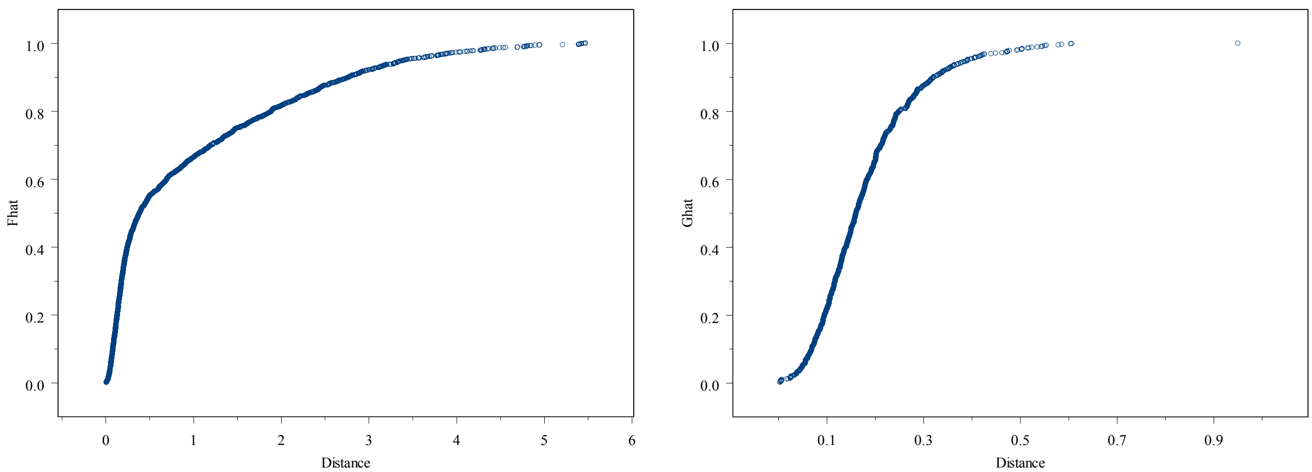


Figure 17. F-hat and G-hat analyses for the study of deviation from CSR for the distribution of sawmills in the U.S. South for the year 2005.

Khat and Lhat analyses are used to further explore deviations from a homogenous arrangement of data points (Figure 19) while comparing them to a generated dataset that follows a homogenous spatial distribution. The empirical distribution of sawmills in the U.S. South is compared against a set of simulated points that follow a spatial random process. Figure 18 presents the results of both analyses for the data corresponding to the U.S. South region. The circles in blue represent the empirical spatial distribution of points while the brown lines denote a set of generated points following a random spatial process. Deviation of data from complete spatial randomness corroborates the findings of the Fhat and Ghat tests.

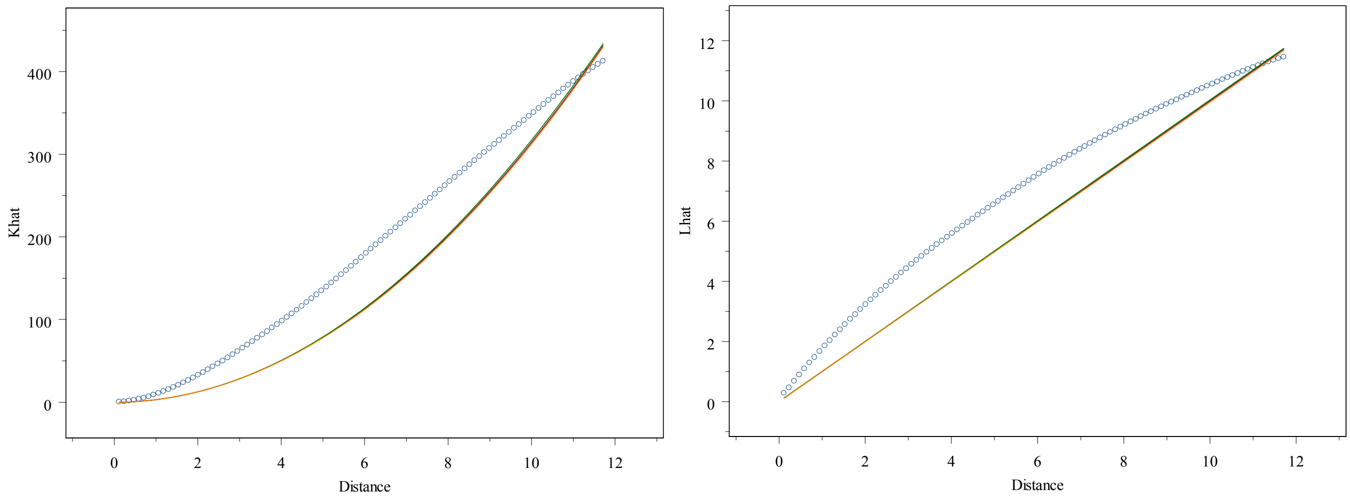


Figure 18. K-hat and L-hat analyses for the study of deviation from CSR for the distribution of sawmills in the U.S. South for the year 2005.

7.3 Deviations from Complete Spatial Randomness in Spatial Distribution of Sawmills in the U.S. Texas Region

The same procedure to determine deviations from complete spatial randomness was performed to the data corresponding to sawmills in the U.S. Texas region. A total of 124 sawmills were identified in the region by the USDA Forest Service Southern Research Station (2005a). Figure 19 shows the distribution of sawmills where each enterprise is identified with a circle. Data is projected in geographic degrees, the longitude coordinates extend from -101.912 to -93.697 degrees, and the latitude coordinates extend from 27.511 to 33.640 degrees. It is evident from the scatter plot that sawmills are not spread across the state but tend to cluster. The majority of sawmills in the state of Texas locate on the eastern part of the state close to the Louisiana border.

Figure 20 presents the Fhat and Ghat analyses for the study of point data distribution in the Texas region. Both analyses suggest deviation from complete spatial randomness in the spatial arrangement of sawmills in the State.

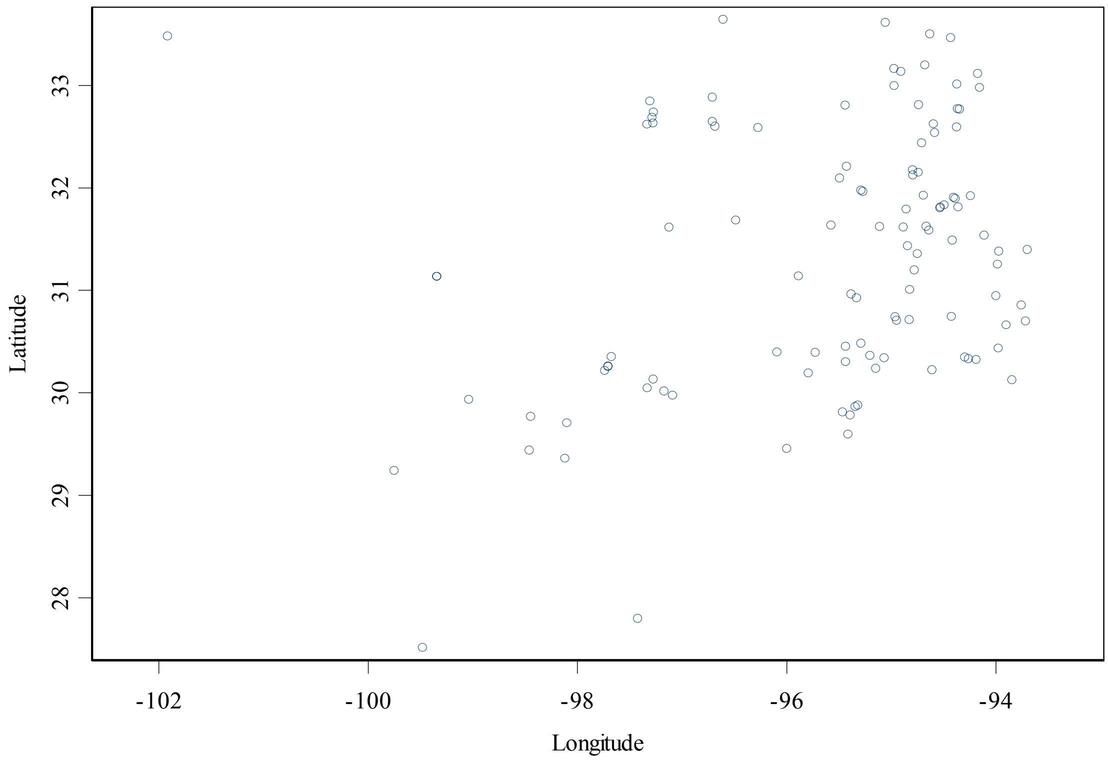
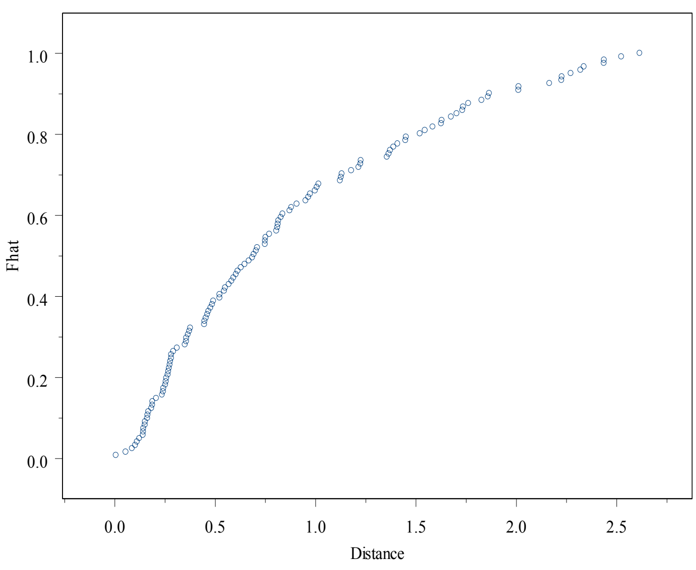


Figure 19. Scatter plot of Sawmills in the U.S. Texas region for the year 2005.



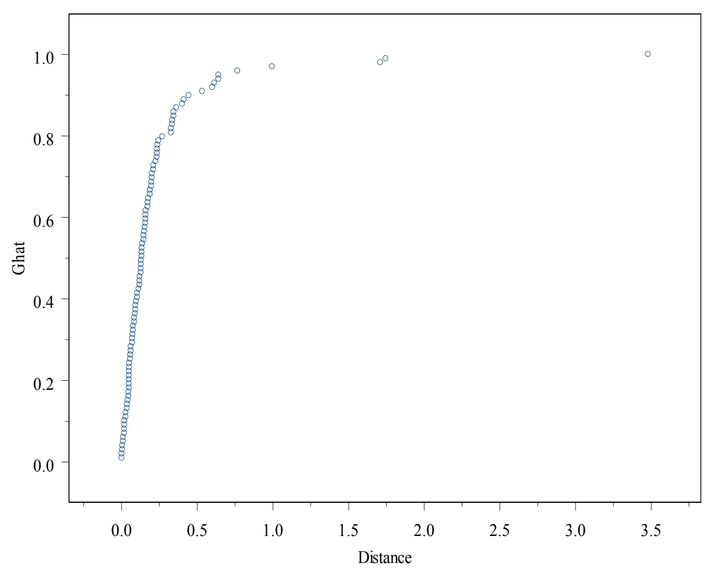


Figure 20. F-hat and G-hat analyses for the study of deviation from CSR for the distribution of sawmills in the U.S. Texas region for the year 2005.

An excess number of long distance neighbors in the Fhat analysis and a large number of points at short neighbor distances in the Ghat analysis provide evidence of clustering of data points. Khat and Lhat analyses are used to further explore deviations from complete spatial randomness (Figure 21). The empirical distribution of the data is compared against a set of simulated points that follow a spatial random process. Figure 21 presents the results of both analyses for the data corresponding to the U.S. Texas region. The dots in blue color show the empirical spatial distribution of points while the brown lines denote sets of generated points following a homogenous spatial distribution process.

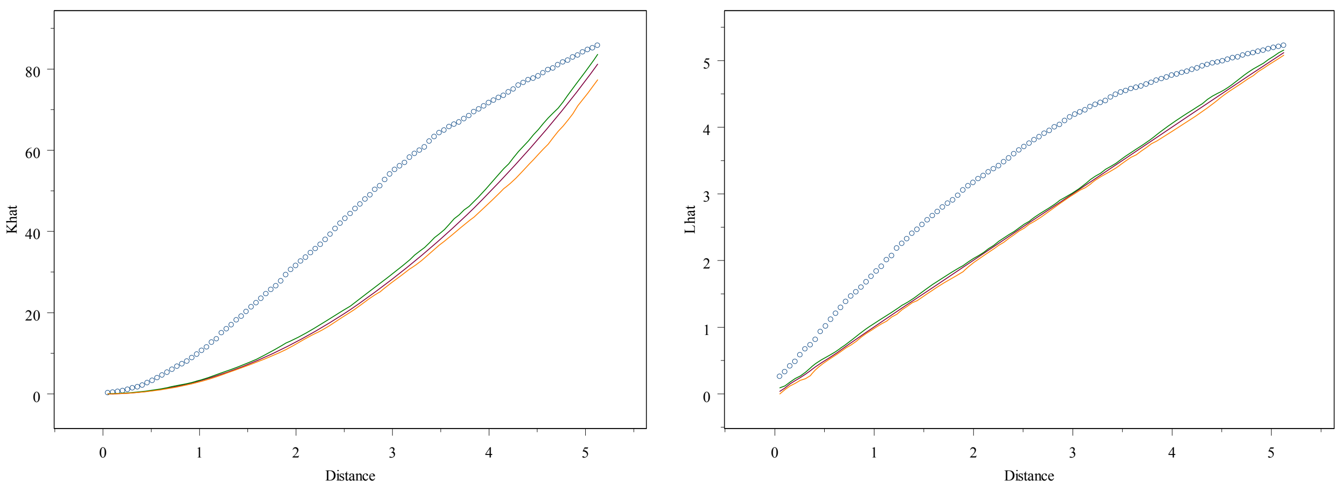


Figure 21. K-hat and L-hat analyses for the study of deviation from CSR for the distribution of sawmills in the U.S. Texas region for the year 2005.

A formal hypothesis test was carried out to determine whether the deviation from spatial randomness is statistically significant. A Chi-square test for which the null hypothesis is spatial random distribution of observations was rejected. To carry out the test statistic the 1,910 sawmills in the South and Texas region were divided in five categories corresponding to the number of sawmills per county (Table 54). Following a Poisson distribution, a Chi-square test for spatial randomness with four degrees of freedom equals 764.45. This test-statistic has a p-

value much less than 0.0001 providing strong evidence against homogenous distribution of sawmills in the U.S. South and Texas regions.

Table 55. Number of sawmills per county and corresponding frequencies in the U.S. South and Texas regions.

Number of sawmills per county	Frequency
0	549
1	235
2	139
3	80
>4	141
Total	1,144

The exploratory study of the location of sawmills in the U.S. South and Texas regions provides strong evidence that the industry is not homogenously distributed and rather suggests a clustering pattern. This finding consists in a favorable justification to the formal inclusion of a spatial dimension in a model for the incidence of softwood lumber industry. Clustering in the data suggests that the assumption of independence between observations may be violated as the occurrence of industry in one county makes it more likely to have industry in surrounding counties as well. As sawmills tend to concentrate in space, the probability of having a county hosting the lumber industry may be affected by the occurrence of industrial activity in nearby counties. This spatial effect may be captured in the form of a spatial autoregressive functional form presented in the next section. As in a time series autoregressive model observations are a function of previous time period, the incidence of industry in one county is a function of industry in surrounding counties. The main difference in the latter is that the effect is multidirectional. The relative weight of each county in the region under study on an observation is given by a spatial contiguity matrix as discussed in the Methods section.

CHAPTER 8. RESULTS: SPATIAL GEOGRAPHICAL ANALYSIS AS A TOOL FOR ANALYSIS OF A RESOURCE-BASED INDUSTRY

Before getting into the formal spatial analysis of the location of sawmills in the U.S. South it is important to stress the distinction between two approaches traditionally followed when working with a dataset with a spatial dimension (Anselin 1988). The first one, usually followed by econometricians, starts from a theoretical framework that imposes a spatial dependence structure a priori. The theory-based structure is incorporated in a model for statistical analysis. The second approach uses the data itself to make inferences about the appropriate form of spatial dependence based on indicators of data correlation such as variograms or covariograms. This second method is usually followed by statisticians and is not the main approach taken in this dissertation. Nevertheless, the exploratory analysis of data provides support to the model specification and the formal incorporation of spatial autocorrelation. Tools from both approaches are used in this section.

Econometric analysis is performed at the county level, each county identified by its unique FIPS code. The coordinate information for each county is determined as the latitude and longitude corresponding to the county's centroid. This information is obtained using software package GeoDa, developed at the Spatial Analysis Laboratory at the University of Illinois (Anselin et al. 2005). The total number of counties included in the study is 1,144. This figure includes all counties in the U.S. South, except for Kentucky. The reason for the not inclusion of Kentucky is because of lack of information available at the time of the study for log prices. Texas has a total 356 counties. The total number of sawmills in the U.S. South and Texas is then 1,580. The distribution per state is presented in percentage form in Figure 22.

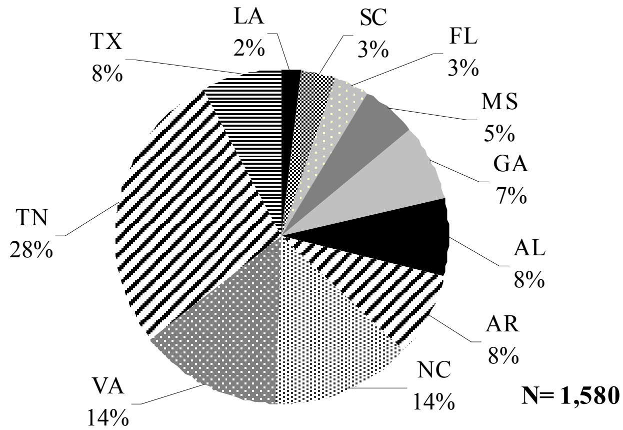


Figure 22. Distribution of sawmills per states included in the spatial analysis of the U.S. Lumber industry.
Source: USDA, Forest Service Southern Research Station (2005a).

The number of sawmills were aggregated at the county level. A histogram of the distribution of the number of mills observed per county is presented in Figure 23. This figure shows the frequency of the number of sawmills found per county. Notice a large frequency of counties with no sawmills corresponding to the value of 0 that results in a skewed distribution.

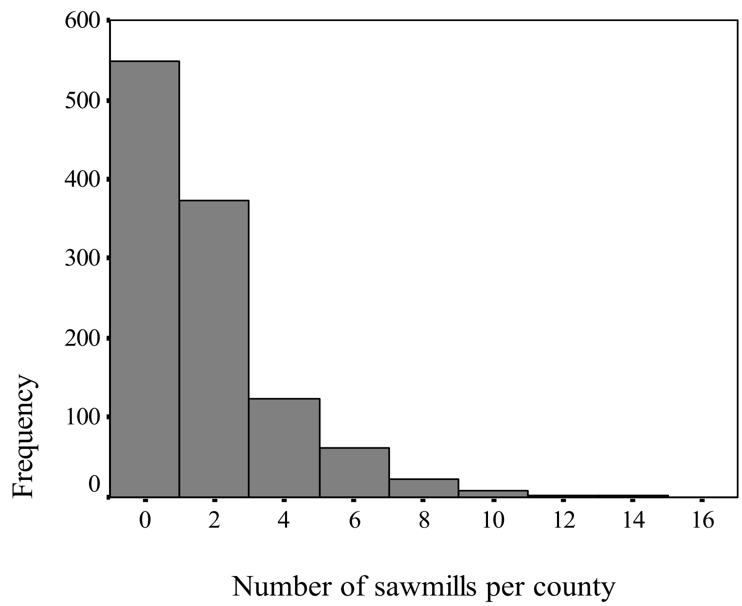


Figure 23. Histogram of the frequency of the number of sawmills per county (n=1144).

The spatial model for the presence of the softwood lumber industry per county is then expressed as the probability that a county hosts the lumber industry (y=1) as a function of the five principal factors identified in the common factor analysis. This relationship is given by $P(y=1) = f(R,P,H,C,A,E)$, where R represents the industry primary input resource, P indicates policies, education, taxes and services, H accounts for human resources, C indicates competition within the industry, A captures variables regarding market accessibility, and E represents energy and land costs. Table 55 shows a list of the variables used as proxies for these factors. The first variable corresponds to the presence of the softwood lumber industry in a county which is used as the dependent variable. The price of logs (dollars per ton) represent cost of the main raw material, and variables for total woodland and land under Conservation Reserve Programs are used for proxies for the local availability of the resource. State-specific dummy variables are used to capture specific policies and taxes. In the analysis the state of Texas is left out as the base level. The presence of a Society of American Forester (SAF) certified program in a county is identified as a 1 or 0 whether there is a formal SAF accredited offered by a university in a county or not. Various variables are included as proxies for Human Resources (H) including total population, total number of unemployed peopled, number of high school graduates 25 years and older, number of college graduates 25 years and older all for the year 2000, and personal income. However, only unemployment is kept in the model for analysis because of the strong correlation between them (linear correlation higher than 0.9). Inclusion of all these variables would result in the introduction of severe multi-collinearity in the model. Unemployment is kept because it is considered the variable that best reflects the availability of labor in a county. Several distances from ports and town are generated to capture the relation between the proximity of sawmills and

access to markets. Also, the presence/non-presence (1/0) of a highway in a county is used as a proxy for accessibility to markets (A).

Table 56. Variables, format and sources of data for the spatial analysis for the Softwood Lumber Industry.

Proxy and Period of Content	Variable Name	Factor	File	Source/Publication
U.S. Wood-Using Mill Locations in 2005	SAWMILL	N/A	SHP	Southern Station, USDA Forest Service/2005
Average stumpage price in first quarter of 2000	PRICE_LOG_D	R	CSV	Timber Mart-South Notes/2006
Total woodland, total farms in 2002	WOOD_FARM	R	CSV	2002 Census Of Agriculture/2006
Land under Conservation Reserve or Wetlands Reserve Programs (Total farms) in 2002	CRP	R	CSV	2002 Census Of Agriculture/2006
Education – University offering Forestry/Forest Products degree in 2006	SAF	P	CSV	Society of American Foresters/ND
Dummy variable generated per each state to capture state specific policies and taxes.	STATE DUMMIES	P	N/A	Profiles of America - ERS/USDA/ 2004
Total Population in 2000	TOTP	H	CSV	Profiles of America - ERS/USDA/ 2004
Unemployment in 2000	UNEMP_2000	H	CSV	Profiles of America - ERS/USDA/2004
Number High School Graduates 25 Years and Older in 2000	HSGRAD2000	H	CSV	Profiles of America - ERS/USDA/2004
Number of College Graduates 25 Years and Older in 2000	COLGRADS_2000	H	CSV	Profiles of America - ERS/USDA/ 2004
Personal Income ($) in 2001	PI_2001	H	CSV	Profiles of America - ERS/USDA/ 2004
10 miles distances from port (date is not determined)	PORT_10	A	SHP	U.S. Department of Transportation/1998
25 miles distances from port (date is not determined)	PORT_25	A	SHP	U.S. Department of Transportation/1998
50 miles distances from port (date is not determined)	PORT_50	A	SHP	U.S. Department of Transportation/1998
50 miles distance from town with a population of at least 50,000 (date is not determined)	POP_50	A	SHP	National Transportation Atlas Database/2006
50 miles distance from town with a population of at least 100,000 (date is not determined)	POP_100	A	SHP	National Transportation Atlas Database/2006
Presence of a highway in 1999	HWY	A	SHP	NationalAtlas.gov/2001
Average industrial price (cents per kilowatt/hour) per State in 2004	ELECT	E	CSV	Energy Information Administration/ 2005
Median house value ($1000) in 2000	HOUSE_V_THOU	E	CSV	U.S. Census 2000/ND
Sales of forest products, excluding Christmas trees and maple products ($1000) in 2002	SALES_FP	G	CSV	2002 Census Of Agriculture/2006
Average Annual Precipitation above 3000 mm. over1961-1990	HIGH_RAIN	G	SHP	NationalAtlas.gov/2001

*CSV (Comma Separated Value Sheet), SHP (Shapefiles).

A spatial weight matrix attempts to capture competition effects from industry in surrounding counties. The cost of electricity as dollars per kilowatt per hour is used as a proxy for the cost of energy (E). The median value of houses in a county is used as a proxy for cost of land. Other variables are included to account for the geographical coincidence of sawmills and sales of forest products (SALES_FP), and a high rainfall level (higher than 3,000 mm) is used as a proxy for ecological variables.

A conventional Probit regression model using maximum likelihood was estimated for what I call a "Full model". This Full model includes all explanatory variables corresponding to the original list presented in Table 56 after discarding variables that would introduce a high degree of multi-collinearity in the model. Explanatory variables in the Full model are presented in Table 57. The coefficients for the explanatory variables in the model have expected signs. There is an inverse and statistically significant (α=0.05) relationship between the likelihood of observing a county host lumber industry with prices for logs, cost of energy and house values. Other variables with a negative sign are proximity to a town with a population of at least 50,000 or a port, the presence of a SAF accredited program and land under Conservation Reserve Programs. None of the latter is statistically significant. The presence of a highway, the geographic coincidence of the industry with sales of forest products and the availability of woodlands are all statistically significant and have a direct relationship with the dependent variable. The number of people unemployed that is used as a proxy for labor availability is marginally significant with a p-value of 0.06. The dummy variable capturing state-specific effects for North Carolina is the only one significant at α=0.05 and has a positive effect.

A reduced model that includes only the variables identified as significant (α=0.05) using a Probit model assuming no spatial correlation is used as the starting point in the next section.

This Reduced model is used to be compared against the Full model and determine the significance of the spatial effects as more variables are added to the model.

Table 57. Probit regression model coefficients using maximum likelihood estimation and no spatial correction.

Variable	Coefficient	Std Deviation	z	p-value
CONSTANT	2.6958	0.786	3.43	0.001
PRICE_LOG_D	-0.0239	0.004	-6.22	<0.001
UNEMP_THOU	0.03182	0.017	1.88	0.060
MAX_HWY_1	0.38625	0.147	2.62	0.009
MAX_POP_50	-0.0202	0.123	-0.16	0.870
MAX_PORT_5	-0.0087	0.102	-0.08	0.933
HIGH_RAIN	0.18753	0.125	1.50	0.134
SALES_FP	0.00078	0.000	3.10	0.002
MAX_SAF	-0.2313	0.274	-0.84	0.399
SUM_CRP_V	-0.0003	0.000	-0.68	0.495
SUM_WOODFA	0.00068	0.000	5.51	<0.001
ELECT	-0.0042	0.001	-3.07	0.002
HOUSE_V_TH	-0.0056	0.002	-2.99	0.003
AR	-0.2366	0.259	-0.91	0.361
AL	0.08311	0.287	0.29	0.772
FL	0.41057	0.239	1.72	0.085
GA	-0.1866	0.200	-0.93	0.350
LA	-0.1088	0.233	-0.47	0.640
MS	-0.1056	0.233	-0.45	0.651
NC	0.71188	0.198	3.60	0.001
TN	0.08592	0.283	0.30	0.762
VA	-0.0699	0.233	-0.30	0.764

Building on the work of McMillen (2003) more variables with a geographic dimension are included in the regression model are included in an attempt to determine changes in the strength of the spatial process. It is expected that the strength of the process will decline as more explanatory variables are added. McMillen (2003) argues that the presence of a spatially correlated effect is partly given by model misspecification.

8.1 Tests for Spatial Correlation

The next step in the spatial analysis of the industry involves the formal specification of a spatial dimension in the model. This can be given by a spatial autoregressive specification, a spatial correlation in the residuals, or both. The formal inclusion of spatial correlation is also suggested by using various tests for the presence of a spatial effect. These tests are performed to

what I have identified as the "Reduced" and the "Full" models. The reduced model only includes as explanatory variables the predictors identified as statistically significant following a conventional Probit specification. The Full model makes use of all explanatory variables presented in Table 57. The results of the spatial correlation tests for both models help on exploring the strength of the spatial process.

Results of the tests for the reduced model are presented in Table 58. Based on the results from a Lagrange Multiplier, Likelihood and Wald tests for the reduced model there is strong evidence of spatial correlation in the residuals. A Moran I-test for correlation in the residuals was also fitted and resulted in a test statistic of 3.0 and an associated p-value of 0.002 rejecting the null hypothesis of no spatial correlation.

Table 58. Tests for spatial correlation in residuals (Reduced model).

	Lagrange Multiplier	Likelihood ratio	Wald
Value	7.33703622	6.90510426	12.03673947
Chi-square (1) at α=0.01	6.63500000	6.63500000	6.63500000
Marginal probability	0.00675481	0.00859500	0.00052162

The results for a spatial autoregressive model using a Lagrange Multiplier test reject the null hypothesis of no spatial correlation in the residuals. The test LM test-statistic is 8.261 and a p-value of 0.004. This result suggests that the spatial autoregressive model still results in spatial correlation in the residuals.

Results of the tests for the Full model are presented in Table 59. Based on the results from a Lagrange Multiplier, Likelihood and Wald tests for the reduced model there is weaker evidence of spatial correlation in the residuals. A Moran I-test for correlation in the residuals was also fitted and resulted in a test statistic of 2.43249618 and an associated p-value of 0.01499515. Although the results of the Moran I-test suggest presence of spatial correlation in the residuals, this is considerably smaller relative to the reduced model suggesting that the strength of the

process is reduced as more variables, as it is the case of the state-specific dummies, are included in the model.

Table 59. Tests for spatial correlation in residuals (Full model).

	Lagrange Multiplier	Likelihood ratio	Wald
Value	2.67182695	2.72187883	3.01107049
Chi-square (1) at α=0.01	6.63500000	6.63500000	6.63500000
Marginal probability	0.10213872	0.09898144	0.08269766

The results for a spatial autoregressive model using a Lagrange Multiplier test reject the null hypothesis of no spatial correlation in the residuals. The test LM test-statistic is 9.859 and has a p-value of 0.001. This result suggests that the spatial autoregressive model still results in spatial correlation in the residuals.

The previous tests are used as a guide to guide the specification for the spatial process in the model. The results of the Probit model Bayesian estimation and the effects of correlation in the autoregressive specification or correlated errors further discussed under the reduced and full models are presented in the next section.

8.2 Model Estimation and Incorporation of a Spatially Correlated Specification

The results of a Bayesian Spatial Autoregressive Probit model for the reduced form are presented in Table 60. The pseudo R-squared for the model is 0.6803 for 1,144 observations and 9 explanatory variables. As expected, the coefficients have the same signs as obtained using a conventional probit model with no specification for spatial correlation. The values of the coefficients are also very close to those estimated using maximum likelihood. The absolute value of ρ that denotes the strength of the autoregressive process is 0.082868. Although it is statistically significant from “0” its spatial effect can be considered weak.

When a full model is fitted including all explanatory variables in the Bayesian estimation the R-squared value is slightly increased to 0.6870 which is not surprising as the variables that are statistically significant account for most of the variability in the model. The results are

presented in Table 61. There are no major changes in the values of the coefficients; all signs remain the same in the Full model specification. Regarding the strength of the autoregressive process, the value of ρ is reduced to 0.056648 which is even weaker although still statistically significant.

Table 60. Bayesian spatial autoregressive Probit model posterior estimates (reduced model).

Variable	Coefficient	Std. Error	z	p-value
CONSTANT	2.250525	0.550	4.090	<0.001
PRICE_LOG_D	-0.026617	0.004	-6.597	<0.001
UNEMP_THOU	0.037148	0.023	1.599	0.038
HWY_1	0.493961	0.189	2.608	0.005
SALES_FP	0.001041	0.000	3.117	<0.001
WOODFA	0.000705	0.000	5.261	<0.001
ELECT	-0.003240	0.001	-3.703	<0.001
HOUSE_V	-0.005901	0.002	-2.702	0.002
NC	0.771239	0.207	3.717	<0.001
ρ	0.082868	0.053	1.575	0.008

Table 61. Bayesian spatial autoregressive Probit model posterior estimates (full model).

Variable	Coefficient	Std. Error	z	p-value
CONSTANT	3.225944	1.0714	3.0110	<0.001
PRICE_LOG_D	-0.026164	0.0050	-5.2088	<0.001
UNEMP_THOU	0.031605	0.0210	1.5039	0.051
MAX_HWY_1	0.484764	0.2049	2.3657	0.011
MAX_POP_50	-0.029693	0.1459	-0.2035	0.406
MX_PORT_5	0.020857	0.1293	0.1613	0.435
HIGH_RAIN	0.246999	0.1606	1.5376	0.055
SALES_FP	0.001055	0.0004	2.9635	<0.001
SAF	-0.180208	0.3670	-0.4911	0.291
CRP	-0.000402	0.0006	-0.6860	0.253
WOODLAND	0.000782	0.0002	4.5465	<0.001
ELECT	-0.005096	0.0019	-2.7531	0.002
HOUSE_V	-0.007104	0.0024	-2.9368	0.001
AR	-0.298847	0.3577	-0.8355	0.204
AL	0.023633	0.3813	0.0620	0.448
FL	0.427331	0.3142	1.3602	0.081
GA	-0.312652	0.2671	-1.1706	0.116
LA	-0.221505	0.3077	-0.7198	0.250
MS	-0.223767	0.2877	-0.7777	0.216
NC	0.780321	0.2652	2.9422	0.001
TN	0.050424	0.3556	0.1418	0.428
VA	-0.134501	0.2959	-0.4546	0.331
ρ	0.056648	0.0435	1.3032	0.015

A spatial error specification is also fitted to both Reduced and Full models. The results for the Bayesian spatially correlated errors specification of the reduced model are presented in Table 62. The values and level of significance of all coefficients is similar to the previous models. The value for the coefficient λ that captures the strength of the spatial correlation in the errors has a higher value than the spatial autoregressive model and it is statistically significant. Nevertheless, the strength of the process is still not of major consideration.

Table 62. Bayesian Spatial Correlated Error Probit Model (reduced model).

Variable	Coefficient	Std. Error	z	p-value
CONSTANT	2.170418	0.550	3.944	<0.001
PRICE_LOG_D	-0.023802	0.004	-5.899	<0.001
UNEMP_THOU	0.030340	0.023	1.306	0.051
HWY_1	0.421166	0.189	2.223	0.013
SALES_FP	0.000888	0.000	2.659	<0.001
WOODFA	0.000668	0.000	4.985	<0.001
ELECT	-0.003151	0.001	-3.601	<0.001
HOUSE_V	-0.005559	0.002	-2.545	0.001
NC	0.733476	0.207	3.535	<0.001
λ	0.112259	0.053	2.134	<0.001

When estimating the Full model, the value of λ declines slightly but stays statistically significant suggesting the presence of a spatial process in the data. This is an expected effect because as discussed by Anselin (2001) the integration of data from various sources and scales tends to result in spatially dependent process.

Autocorrelation is a common problem in spatial data, and significant advances have been made in devising parametric models that account for it. McMillen (2003) stresses that autocorrelation is often produced spuriously by model misspecification. McMillen (2003) consider that spatial correlation may be the result of incorrect functional form or a problem of missing variables that are correlated over space. Supplementing an incorrectly specified model with variables account for spatial variability, such as dummy variables for regions and interaction terms, would produce a more accurate model specification.

Table 63. Bayesian Spatial Correlated Error Probit Model (full model).

Variable	Coefficient	Std Deviation	z	p-value
CONSTANT	2.718172	1.0098	2.6917	0.004
PRICE_LOG_D	-0.023821	0.0040	-5.9434	<0.001
ELECT	-0.004233	0.0017	-2.4244	0.009
UNEMP_THOU	0.027878	0.0189	1.4735	0.063
MAX_HWY_1	0.401598	0.1774	2.2635	0.013
MAX_POP_50	0.001425	0.1495	0.0095	0.506
MX_PORT_5	-0.040100	0.1265	-0.3169	0.362
HIGH_RAIN	0.191976	0.1511	1.2702	0.105
SALES_FP	0.000841	0.0003	2.6118	0.003
SAF	-0.177176	0.3530	-0.5018	0.300
CRP	-0.000140	0.0005	-0.2834	0.391
WOODLAND	0.000687	0.0001	4.7055	0.000
AR	-0.183327	0.3329	-0.5506	0.294
AL	0.158857	0.3827	0.4151	0.337
FL	0.463840	0.2736	1.6953	0.045
GA	-0.211427	0.2596	-0.8143	0.215
LA	-0.202261	0.2837	-0.7129	0.243
MS	-0.139096	0.2897	-0.4802	0.313
NC	0.758847	0.2387	3.1795	0.001
TN	0.092146	0.3306	0.2787	0.404
VA	-0.040944	0.3056	-0.1340	0.445
HOUSE_V	-0.006116	0.0023	-2.6931	<0.001
λ	0.109937	0.0579	1.8976	<0.001

8.3 Predicted Potential Developments in the Softwood Lumber Industry in the U.S. South

The next step in the analysis is to estimate the predicted probability of a county hosting the softwood lumber industry and compared it to current status. Estimated probabilities were calculated for all 1,144 counties in the study and compared to the current value of the dependent variable. The value of observed industry presence was subtracted from the estimated probability of industry presence creating a new variable YDIFF. When the absolute value of YDIFF is larger than 0.80 that county is identified as an area that could host potential new developments in the industry. The value of 0.80 is set as a minimum threshold for the strength of the prediction versus the non-presence of the softwood lumber industry in that county.

Table 64 shows the results for the non-spatial Probit model using maximum likelihood estimation and a Probit model that incorporates a spatial dimension in the model. For estimation of the spatial Probit model an autoregressive specification was used. This one specification was

used because of the previous finding of industry clustering which suggests that sawmills tend to locate in areas where other sawmills are. This spatial distribution is directly linked to the availability of the forest resource. Comparing the two models, the one with a spatial correlation autoregressive process suggests that 25 counties could potential host new developments in the sector, compared to 23 estimates using the non-spatial probit model. It also predicts that further developments could occur in Alabama, Arkansas and Mississippi compared to the non-spatial specification, but it also predicts that five, instead of six, counties could host development in North Carolina.

Table 64. Number of counties and states where new developments in the Softwood lumber industry in the U.S. South region could occur.

State	Non-spatial Probit model	Spatial Probit model
Alabama	3	4
Arkansas	5	6
Mississippi	0	1
North Carolina	6	5
South Carolina	1	1
Tennessee	5	5
Texas	2	2
Virginia	1	1
TOTAL	23	25

The actual location of the counties where new developments could be expected to occur is depicted in Figures 25 and 26. Figure 24 depicts the counties where new developments in the softwood lumber industry are estimated to occur using a conventional probit model with no-spatial correlation.

Figure 25 shows the counties where new developments in the softwood lumber industry are estimated to occur using a Probit model with a spatial correlation dimension. Most of future developments are expected to take place in Arkansas, North Carolina and Tennessee. Notice that the predictions of both models is very similar and actual changes are due to rounding and the

0.80 threshold used as a selection of counties that are most likely to have developments in the softwood lumber industry.

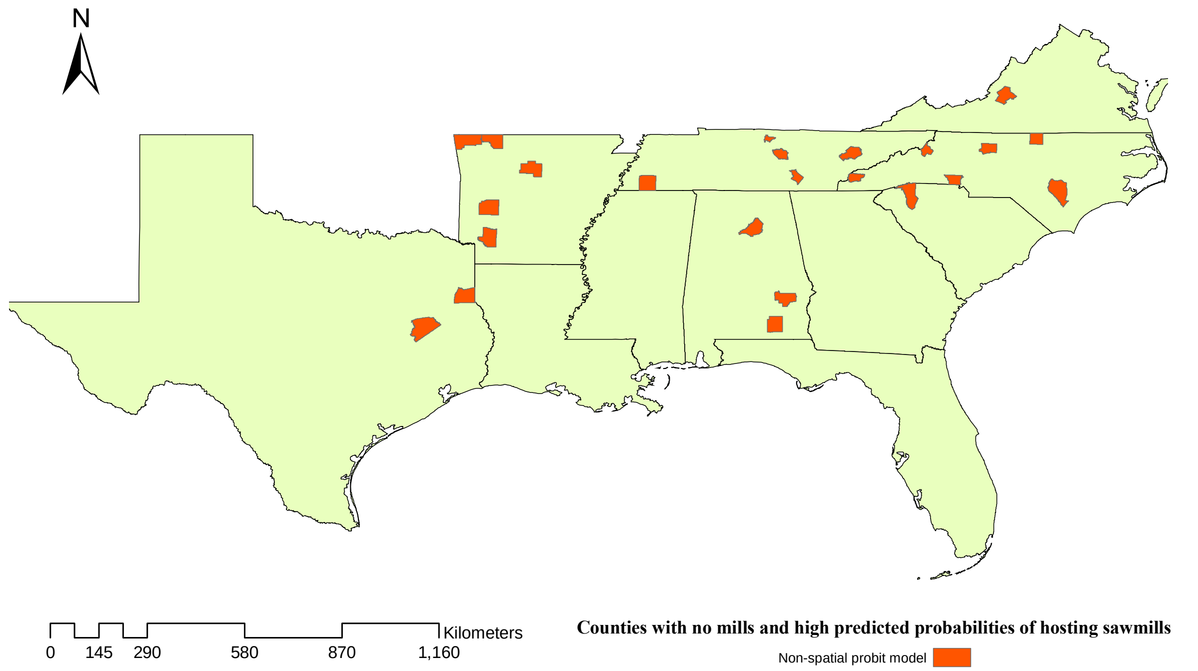


Figure 24. Predicted new developments in the Softwood Lumber Industry in the U.S. South as estimated using a conventional Probit model with no-spatial correlation.

It is important to stress that these predictions are based on the current values for the variables used in the model. Changes in such variables (i.e. new levels of unemployment, log prices or land values) will result in different estimates. Furthermore, this model specification assumes that all other variables not included in the model stay constant.

It is also important to consider that these results are determined based on a regional study and specific sites for location require further assessment. The model developed indicates where new developments in the industry could occur but the actual location of a new sawmill demands the detail study of the county. After a county has been identified as an area of potential development, specific tools for identification of optimal location such as in McCauley and

Caulfield (1990) or other GIS based approaches can be used to help identify a more specific site for a new mill.

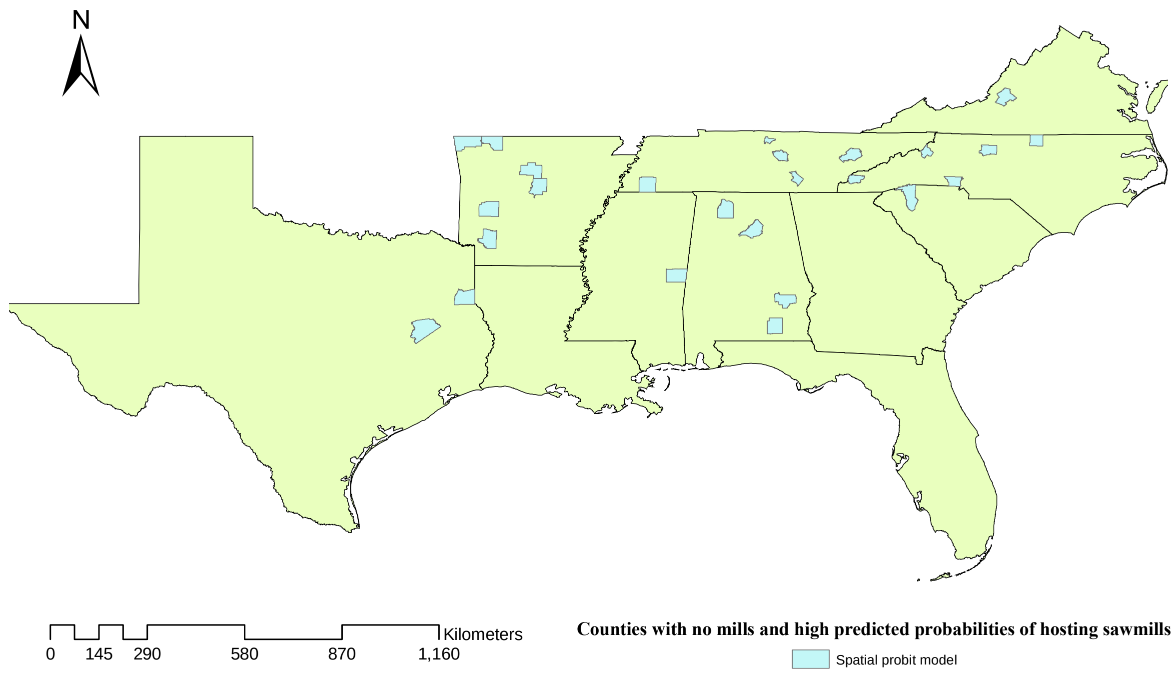


Figure 25. Predicted new developments in the Softwood Lumber Industry in the U.S. South as estimated using a Spatial Autoregressive Probit Model.

CHAPTER 9. RESULTS: DEVELOPMENT OF THE LUMBER INDUSTRY SECTOR BETWEEN 1999 AND 2005 IN THE U.S. SOUTH

Helburn (1943) considers that the geographic pattern of an industry may remain unchanged over time because of inertia. It may be more profitable for a firm to simply remain in the current non-optimal location because the cost of abandoning an old plant and building a new one may be prohibitively high. In industries where fixed capital is a considerably large proportion of total costs it can be expected to see no movement at all in terms of shifting industry from one region to another. Industrial migration rarely takes the form of actual plant relocation. New firms entering the market rather try to identify better locations that may give them a competitive edge. Then, as it is predicted by industrial theory (Helburn 1943), older factories with higher costs or inferior products are simply driven out of business. The lumber industry can experience such a situation and as companies become less efficient and manufacturing equipment turns obsolete, more capacity may be transferred from older smaller to newer larger companies.

In this part of the analysis two cross-sections are taken for 1999 and 2005 to compare how the number of sawmills has changed over time. In this section I also include information from survey participants about their expectations regarding new developments in the softwood lumber industry. Both results are presented in the following subsections.

9.1 Change in the Number of Sawmills per State from 1999 to 2005 in the U.S. South and Texas.

The total number of sawmills was obtained from two cross-sections corresponding to the U.S. South and Texas regions in 1999 and 2005 (Table 64). Overall, both regions have experienced a decline in the number of operating sawmills that in the U.S. South represent a reduction of 14 percent in the number of sawmills. Texas also experienced a reduction in the number of sawmills from 154 in 1999 to 124 in 2005 corresponding to a reduction of 19 percent.

Although the number of sawmills do not reflect the total sawmill capacity of the industry in the region, the shutdown of sawmills suggests a concentration of production in fewer firms. This can be inferred based on the information presented by Spelter (2003). Spelter reports that the sawmill capacity in the U.S. South has actually increased 13.5 percent over the 1995-2002 period thanks to improvements in small log sawing technology and the transfer of capacity from the timber-starved U.S. West.

Table 65. Change in the number of sawmills per state from 1999 to 2005 in the U.S. South and Texas.

State	1999	2005	Difference	Percentage change
AL	148	121	-27	-18
AR	128	127	-1	-1
FL	67	53	-14	-21
GA	144	117	-27	-19
KY	371	317	-54	-15
LA	43	32	-11	-26
MS	84	84	0	0
NC	272	215	-57	-21
OK	9	13	4	44
SC	61	51	-10	-16
TN	494	439	-55	-11
VA	254	217	-37	-15
SOUTH	2075	1786	-289	-14
TX	154	124	-30	-19
SOUTH + TX	**2229**	**1910**	**-319**	**-14**

This increase in capacity may be due to the development of sawmills with a higher capacity as it is reported by Spelter (2003). Looking at a point density plot of sawmills in 1999 (Figure 26) and 2005 (Figure 27) for the U.S. South the density tends to have concentrated over time around the center of the South region. Sawmills may grow larger to take advantage of economies of scale allowing them to reach lower average costs of unit of lumber sawn.

Total lumber output for Southern Yellow Pine, the main species manufactured in the region has actually increased over the same period of time (Figure 28). Again, the reduction of number of sawmills accompanied by an increase of manufacturing capacity are indicators of consolidation

in the industry and that older mills with higher costs or inferior products have been driven out of business

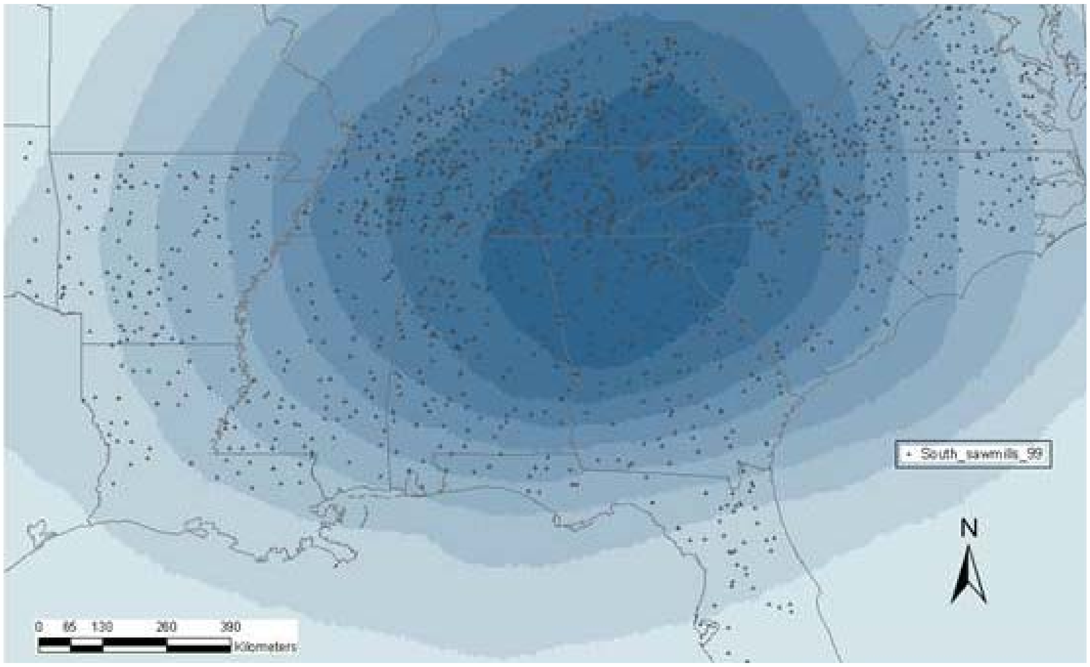


Figure 26. Point density for sawmills in the U.S. South region for the year 1999.
Source: USDA Forest Service, Southern Research Station (2005b).

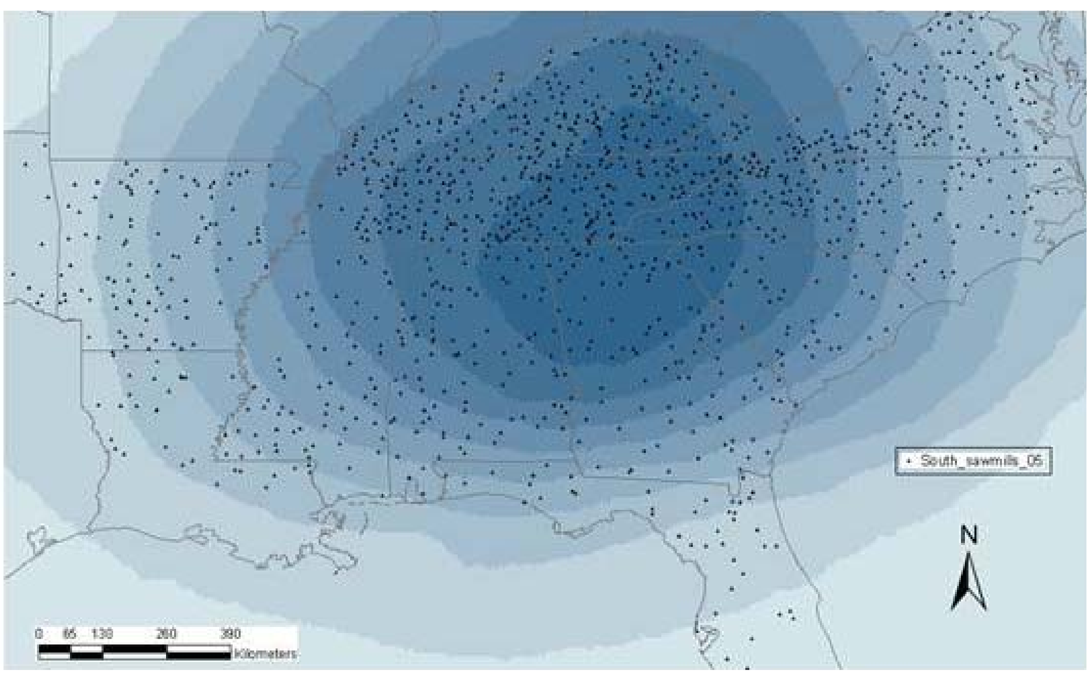


Figure 27. Point density for sawmills in the U.S. South region for the year 2005.
Source: USDA Forest Service, Southern Research Station (2005a).

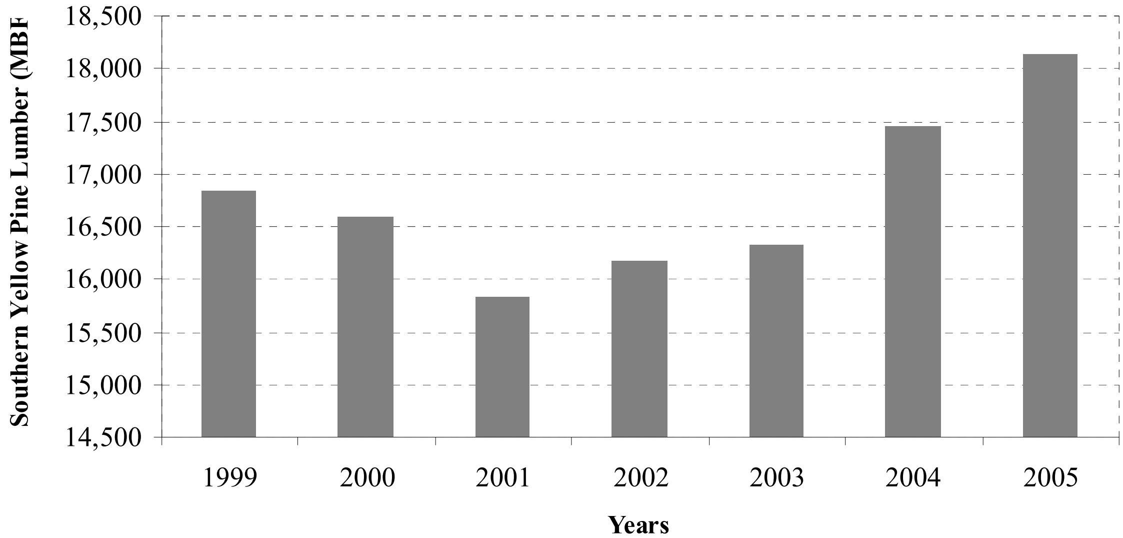


Figure 28. Annual lumber output of Southern Yellow pine from 1999 to 2005 (Million Board Feet).
Source: U.S. Census Bureau (2006a).

Figures 29 and 30 show a point density analysis for sawmills in Texas for the years 1999 and 2005. A total 30 sawmills have been driven out of business and the spatial distribution of the industry has contracted over time in the North eastern part of the State.

Although the total number of sawmills has declined over time its milling capacity for the region has expanded (Spelter 2003). Looking at the point density of the Texas region it is noticeable to see how it has concentrated over a smaller area denoted by a more intense and dark colored region. Sawmills aggregate over this area again, probably taking advantage of lower production costs derived from a larger capacity.

While this part of the analysis looks over the evolution of the industry during the 1999-2005 period it remains a research question to whether if and where any developments in sawmill capacity may occur in the foreseeable future. To address this question, survey participants were asked about their perception of future capacity development in the two major regions in regard to sawmill capacity in the country.

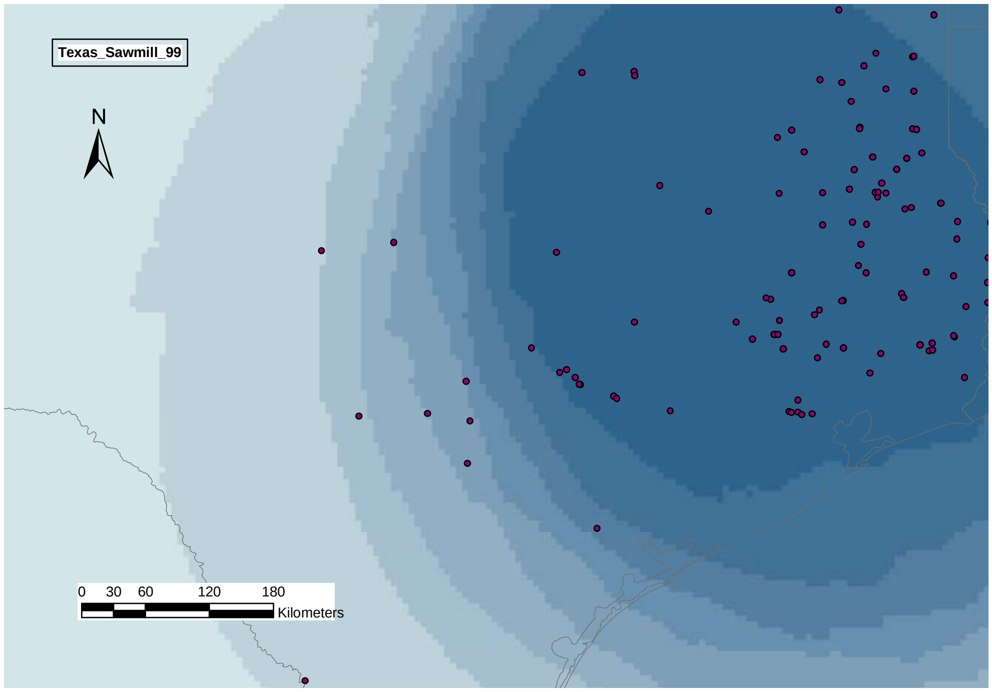


Figure 29. Point density for sawmills in Texas for the year 1999.
Source: USDA Forest Service, Southern Research Station (2005b).

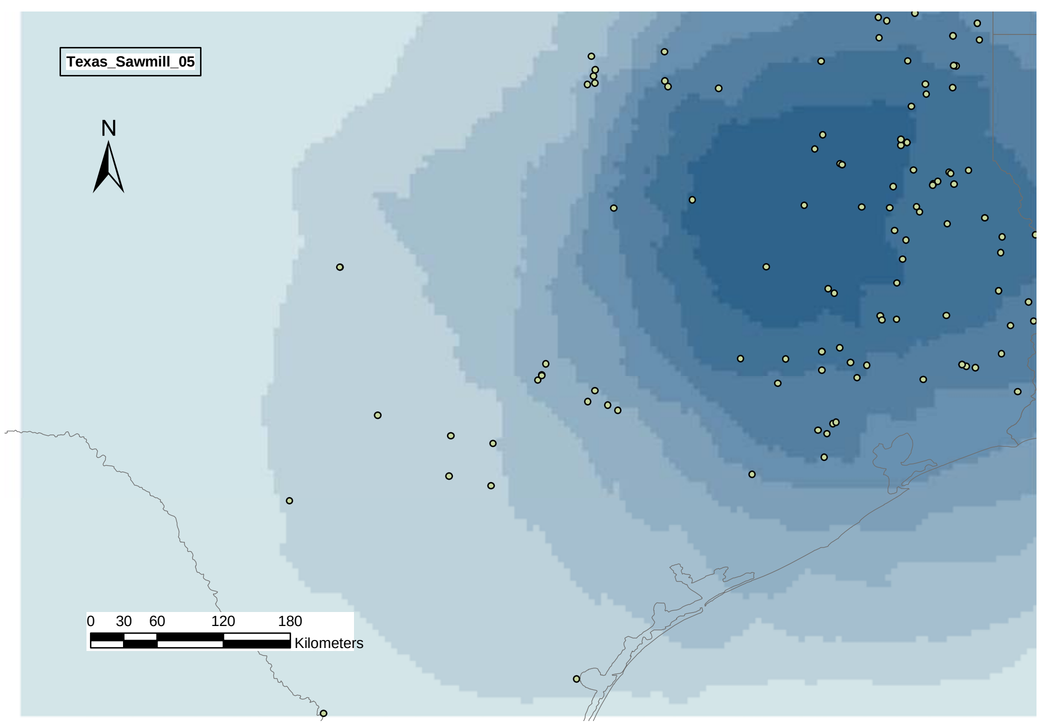


Figure 30. Point density for sawmills in Texas for the year 2005.
Source: USDA Forest Service, Southern Research Station (2005a).

9.2 Expected Developments in Total Sawmill Capacity in the Softwood Lumber Industry.

Study participants were asked for their beliefs regarding future increases in sawmill capacity in different U.S. regions. The majority of respondents consider that the industry will not experience a growth in capacity in the four major regions of the country. Table 66 summarizes the results for potential increase in the U.S. South, West, Northeast and North Central regions.

Table 66. Participants believe in the future increase of sawmill capacity in different U.S. regions.

	Yes	No	Do not know
Do you expect to see an increase in softwood sawmill capacity in the U.S. South in the next 5 years?	19 (23.5%)	40 (49.4%)	22 (27.2%)
Do you expect to see an increase in softwood sawmill capacity in the U.S. West in the next 5 years?	17 (21.0%)	44 (54.3%)	20 (24.7%)
Do you expect to see an increase in softwood sawmill capacity in the Northeast region of the U.S. in the next 5 years?	4 (4.9%)	38 (46.9%)	39 (48.1%)
Do you expect to see an increase in softwood sawmill capacity in the North Central region of the U.S. in the next 5 years?	4 (4.9%)	38 (46.9%)	39 (48.1%)

Although a large percentage of respondents do not consider there will be growth in any of the U.S. regions, perceptions for potential new growth in capacity over the next five years are higher for the U.S. South and West regions. To determine the statistical significance of these responses, a proportion test was carried out. The proportion test compares the observed proportion of responses that consider there will be an increase in sawmill capacity at each region versus a hypothesized proportion of 0.3. This proportion is simply used as a point of reference that considers that 30 percent of respondents would have a positive view of growth in the future. Here, the null hypothesis is that there is at least a 0.3 proportion in favor of each of the four statements. P-values indicate the probability of observation a proportion below the 0.3 level. Results are presented in Table 67.

The test of proportions only failed to reject the hypothesis for the U.S. South region at α=0.05 level. For all other regions the proportion of respondents that expect to see an increase in capacity is statistically lower than the 0.3 hypothesized mean proportion. This finding suggests

that members of the industry may expect to see an increase in sawmill capacity mainly in the U.S. South.

Table 67. Test of proportions for expectations toward an increase in sawmill capacity in the four major U.S. wood products regions.

Variable	Mean proportion	Standard error	t	p-value
Do you expect to see an increase in softwood sawmill capacity in the U.S. South in the next 5 years?	0.235	0.047	-1.285	0.099
Do you expect to see an increase in softwood sawmill capacity in the U.S. West in the next 5 years?	0.210	0.045	-1.770	0.038
Do you expect to see an increase in softwood sawmill capacity in the Northeast region of the U.S. in the next 5 years?	0.049	0.024	-4.922	0.001
Do you expect to see an increase in softwood sawmill capacity in the North Central region of the U.S. in the next 5 years?	0.049	0.024	-4.922	0.001

CHAPTER 10. CONCLUSIONS

The study of factors that influence the location of industries has received attention from the research community from as early as the 1890s with the work of Ross (1896) and captured the interest of other researchers. von Thünen developed a theoretical model for land use based on land rents and product prices that result in the distribution of agricultural uses in concentric circles around a central urban area. Predohl (1928) suggested that the factors determining the location of manufacturing enterprises are those determining specific cost advantages at certain places. Weber (1929) classified industries depending on a material index of production, which refers to the proportion of the weight of used localized input materials to the weight of the manufactured product and determines the location of an industry whether close to raw materials or its final market. Weber based his theory of location of industries on the concept of minimum transportation costs. Hoover (1948) deems that the understanding of how different factors of production are priced helps determining the geographical distribution of industrial activity. Isard (1949) defined the general Theory of Location as one embracing the total spatial array of economic activities, with attention paid to the geographic distribution of inputs and outputs and the geographic variations in prices and costs.

But Krugman (1995b) argues that Location Theory did not achieve major success because of its failure to identify the decision-makers behind industrial location and the lack of data with a spatial dimension that could test their hypotheses. Krugman (1995) argues that different forces attracting and pushing industry away from an urban center result in the spatial distribution of an urban core and a periphery of resource-based industries. Fujita and Krugman (2004) identify these factors as the centripetal and centrifugal forces affecting industry distribution.

Porter (1998abc, 2000) considers that businesses cluster together in geographical areas where competitive advantages can be experienced. Porter (1998b, 2000) identified four factors that determine competitive advantages at a given location which are factor (input) conditions, demand conditions, context for firm strategy and rivalry, and the presence of related and supporting industries. In his study of industries with clustering patterns, Porter (2003) ranks the wood products industry among the top 25 largest clusters in the country based on the number of people employed and spatial concentrations.

The softwood lumber industry is used as a case study for the analysis of factors influencing the spatial distribution of natural resource-based industries because of several factors. There are previous reports (Porter 2003, Braden et al. 1998, Aguilar and Vlosky 2006) that suggest the spatial aggregation of the industry, the development of industry clusters in the U.S., and a spatial dependency of the primary wood products sector. Furthermore, because of the direct link to the forest resource for raw materials in the lumber industry this is a good example of an industry with a definite spatial association.

On the assumptions that the industry can freely locate in the country, information is perfect and available, that decision-makers maximize their utility, and holding everything else constant, a method for analysis is developed. The methods attempt to answer several hypotheses related to location theory, new economic geography, cluster theory, and with practical implications to the development of the industry. A model for the study of decision-makers' preferences and industry behavior is developed to help identify areas for potential new developments in the sector and the country.

A three-component research approach is taken. First, decision-makers in the industry are surveyed to elicit their preferences for location and identify specific attributes that influence the selection of site for current location and hypothetical new developments. Second, the information

is taken to guide the development of a model for industry behavior. Third, two cross sections are studied to examine evolution of the industry over time.

For the first component the two most important regions in the country, as per total output, are selected and the contact information for decision makers is obtained from the premier database to the wood products industry (Random Lengths' Big Book 2006). Owners and managers of sawmills in the U.S. South and Pacific Northwest regions are included in a study of location preferences that follows Dillman (2000) Tailored Design method. Attributes and constructs used in the survey were selected from the literature on location theory and marketing studies and the final survey was pre-tested with members of the Louisiana State University Agricultural Center and a sample of sawmill managers in Louisiana.

Surveys were returned from 21 different states from the two regions almost equally distributed. The adjusted return rate is 18.5 percent and the profile of the average respondent reflects the situation in the specific region and differences between the regions. A Conjoint Analysis for location preferences allows for the estimation of part-worth coefficients for the decision-makers utility and identification of the most important factors affecting location. As expected the cost of the primary resource in a resource-based industry is the most important locational factor. Decision makers identified that cost of logs and distance to the primary input are the most important factors. In a second category other input factors, wages and energy costs, are considered when locating a sawmill. Other variables such as cost of land, the quality of access roads and distance to markets are less important to the production factors. Of all, the distance of a site from final markets is the least important factor. Thanks to developments in road access that results in lower transportation costs, the industry tends to locate near where the primary resource is most available and least expensive while, the location relative to markets is less important. Because responses came from people in different positions, mainly owners and

plant managers, their preferences were compared. It is found that there are statistically significant differences in the weight placed on the importance of raw materials and wage costs when selecting a site to locate a hypothetical new sawmill. Sawmill owners place a higher importance than non-owners, while to non-owners cost of logs is more important than to owners.

Factor Analysis is used to consolidate a long list of attributes into fewer common factors to guide the development a model of industry behavior. Based on the results from the factor analysis six common factors are proposed to determine the current location of sawmills. These factors are named “Policies, regulations & knowledge”, “Human Resources”, “Primary Resource Input”, “Competition”, “Accessibility”, “Energy and other costs”. Prior to formal modeling an exploratory analysis of the spatial arrangement of industries was performed to determine deviations from complete spatial randomness. Several point-distance measures and a Chi-square test provide strong evidence of clustering in the data. Data for this part of the analysis is confined to the U.S. South and Texas regions. These regions were included in the analysis because of the availability of data for the six primary factors influencing location in the lumber industry.

The analysis of the occurrence of the lumber industry in the U.S. South and Texas is carried out at the county level. Each county is identified on whether it hosts industry or not based on data obtained from the USDA Forest Service, Southern Research Station (2005a). The probability of occurrence of industry in a county is modeled as a function of the five location factors and additional variables to account for state specific effects, an ecological variable and a variable for the geographical coincidence of manufacturing and wood product sales. The total sample size is 1,144 observations corresponding to all the counties in the region. A conventional probit model using maximum likelihood estimation and spatially correlated models following Bayesian methods are used for coefficient and standard error estimation. There is no difference

in signs and statistical significance between the models used. Both the spatial autoregressive and spatially correlated error models suggest a weak effect from a spatially correlated process. Various models are fitted and suggest that the spatial correlation between observations decline as more location specific variables are included in the model. Nevertheless, an autoregressive form is used for prediction because of the results of the exploratory analysis suggests that industry tends to happen in geographically concentrated areas, hence the likelihood is higher in areas where industry occurs.

Comparing current industry location versus predicted probabilities, there are potential areas for new developments in the region. The model with no-spatial dimension formally incorporated estimates that 23 counties could host new developments. The spatial autoregressive model provides evidence that a total of 25 counties could experience establishment of new lumber enterprises. The proposed methods can be used as a first step toward the identification of a location for a new sawmill. Once a county has been identified a particular site may be selected based on the methods proposed by McCauley and Caulfield (1990) or a GIS-approach as proposed by Jones et al. (2007)[2].

Although the comparison of two cross sections, 1999 and 2005, indicates the total number of sawmills has declined in the region, industry reports (Spelter 2003) suggest that actual total sawmill capacity has increased. The study of point-density patterns suggests that industry has concentrated over smaller areas in the regions. This finding is congruent with a tendency in the industry to consolidate and take advantages of reported economics of scale present in the lumber industry (Murray 1995). When survey participants were asked about the perceptions on

---

[2] Jones, T.L., Schultz, E.B., Matney, T.G., Grebner, D.L., and D.L. Evans (2007). A forest product/bioenergy mill location and decision support system based on a county-level forest inventory and geo-spatial information. Southern Forest Economics Workshop. San Antonio, TX. March 4-6, 2007.

what regions of the country could experience increases in sawmill capacity in a foreseeable future, the region that was identified as the most promising is the U.S. South, compared to the Pacific Northwest, North Central and the Northeast regions.

10.1 Findings to Research Hypotheses

Regarding the specific hypotheses set to be addressed by this dissertation, these have are responded as follows:

H_1: The Primary Input Material to the Lumber Industry, Logs, is the Most Important Factor Determining the Location of the Industry

There is strong evidence favoring this statement. Respondents' preferences as analyzed in a common factor analysis and a formal conjoint analysis concluded this is the most important factor affecting softwood lumber industry location. The econometric analysis of the industry in the U.S. South and Texas regions further confirm the importance of this factor and predicts developments will occur where input costs conditions are favorable to the industry.

H_2: The Cost of Energy Has a Significant Effect and an Inverse Relation with the Likelihood of Firm Location

The coefficients in the conjoint analysis and the industry behavior model have the expected sign as specified by this hypothesis. As a production input, energy plays an important role to the industry and decision-makers' preferences reflect this situation. The model for industry behavior provides further evidence of the importance of this factor in the organization of the lumber industry.

H_3: Labor costs and Availability Have a Significant Effect on the Choice to Locate a New Softwood Lumber Company

As another production input, the higher cost of labor is associated with a lower preference for location and lower likelihood to observe industry occurrence. These results are corroborated in the Conjoint Analysis and Spatial Econometric model. Labor availability is also

important as there is a strong correlation between labor availability and wage rates. Counties where labor is plentiful and that are associated with lower wage levels could attract the lumber industry.

H_4: Access to Transportation Venues is a Factor that Has a Significant Effect on Attracting Industry

There is no strong evidence that the industry locates near markets or ports, but there is strong evidence that access to primary roads is a key attribute, as captured by the highway variable in the industry econometric model. Since the industry ships out its finalized products to destinations at considerable distances the presence of venues to transport them is of considerable importance. Road transportation as captured by this variable is the most commonly used method of transportation, and thus, the most important.

H_5: As a Resource-based Industry the Softwood Lumber Industry Locates Near the Source of Raw Materials

The lack of significance in the industry behavior model for the market variables and the high statistical significance of the woodland variable are used as proxy to test this hypothesis. The null hypothesis that the proximity to markets is "0" failed to be rejected, while the coefficient for woodlands is positive and highly significant. These findings provide strong evidence supporting this hypothesis. The Conjoint Analysis also confirms this assertion as respondents place a high relative importance on the "Distance to logs" site attribute.

H_6: The Presence of Substantial Final Markets Influences the Location of Softwood Lumber Enterprises

This hypothesis is closely related to the previous one. The variable used to capture proximity of sawmills to final urban markets was not significant. Out of the seven variables included in the conjoint analysis the factor "Distance to Market", although being significant, was the least important of all.

H_7: Land Rent Theory, What is the Effect on Softwood Lumber Enterprises Location?

Although being less important relative to costs of logs, wages and energy, the cost of land plays a significant effect on the location decision of sawmills in the industry. The variable "Median House Value" used as a proxy for costs of land in the spatial econometric model has a negative and significant effect on the probability of observing sawmills in a county. In the lines of the von Thünen tradition and as a centrifugal force in Fujita and Krugman's core-periphery model (2004) higher costs of land associated to urban areas is a force that drives the industry to locate in rural, more remote areas where cost of land per unit area is lower.

H_8: The Presence of University Programs and Research Institutions Has a Significant Effect on Softwood Lumber Industry Location

Jaffe (1986, 1989) and Jaffe et al. (2000) studied the incidence of research investments with the development of new patents. In the softwood lumber industry, location of sawmills near university campus where research in forestry/forest products takes place is not a factor of importance. In a list of 24 factors considered to be of importance to the location of mills, it ranked at the bottom of the list in average mean values. As expected, the effect of this variable is not significantly different from "0" when modeled in the analysis of industry spatial behavior.

H_9: Do Centrifugal and Centripetal Forces in the Krugman and Fujita New Economic Geography Tradition Influence Industry Location?

Industry participants in the study identified the softwood lumber industry to cluster in particular regions of the country. Using a common factor analysis, the factors that influence such spatial arrangement were classified in centrifugal and centripetal forces. The most important forces that promote industry concentration as denoted by Likert scales are the availability of more local suppliers, and greater opportunity to vertically integrate into manufacturing secondary

products. Contrary, increased competition between companies and potential increases in log prices discourage industrial aggregation.

H_{10}: Do Preferences for Location Factors Vary Across Decision Makers?

Comparisons between company owners and respondents holding other positions, the majority being sawmill managers, suggest there are different preferences for the attributes that influence location. Sawmill owners place more importance on the cost related to wages while managers find the cost of raw materials and energy inputs of relative more importance.

10.2 Future Developments in the Application of the Proposed Methods

When asked about future developments in the industry in regards to increased capacity, the U.S. South is identified as the one where new capacity may be built. Compared to other regions, the U.S. South still enjoys access to plentiful forest resources and the conditions for the establishment of wood products industries are favorable. When looking at the results of the softwood lumber industry spatial behavior at least 23 counties where there is no industry currently could experience the arrival of new sawmills. This estimation is made when comparing the observed absence of industry versus the predicted high probability for a county to host the lumber industry.

It is important to consider that these results are determined based on a regional study and specific sites for location require further assessment. The model developed indicates where new developments in the industry could occur but the actual location of a new sawmill demands the detailed study of the county. After a county has been identified as an area of potential development, area specific tools can be used, and geographic information systems could be an important tool to the specific selection of the site such as in McCauley and Caulfield (1990) or Jones et al. (2007).

Because of data availability this study was confined to the U.S. South and Texas regions. Nevertheless, similar research methods can be applied to other regions in the country or other countries in the world to help guide industry development. A spatial dimension is incorporated in a model for industry behavior due to the resource-specific nature of a natural resource-based industry. The analysis of other industry may incorporate the inclusion of a spatial dimension but that remains an empirical question. As experienced in this analysis, a spatial dimension in the form of a formal scalar that captures the strength of spatial autocorrelation of observation, may not be necessary when location specific variables can be incorporated in the explanatory variables of a model. However, this matter also remains an empirical question and should be thoroughly tested.

The methods used in this dissertation constitute a proposed methodology of the assessment of any industry. First, decision makers are enquired about the factors they consider when making a decision, that in this case involves the location of a sawmill, but could be expanded to other areas of research as well as to other industries. Second, a model for industry behavior is built based on the characteristics that have been identified as critical. Depending on the availability of data proxies can be used to capture variables that cannot be observed directly. One of the major reasons for the use of a spatially correlated model is that the information obtained often comes in different levels of scale and result in the spatial correlation of observations. Third, after a model have been refined and a specification that closely represents the research problem, the identification of areas of potential expansion (or contraction) of the industry can be made to help guide future developments.

As it is the case of any model that depends on secondary information, different scales and sources of data it is prompt to measurement errors. The current model has been developed given

the information available at the time. As more detailed information comes available it should be incorporated in the model for further refinement and analysis.

An area of future development is the dynamic analysis of the industry. More specific data on industry capacity could be used to model how an industry "migrates" over time and identify the key factors determining such changes. Further, as a national model can be developed it will provide better insights of the industry and by understanding previous industrial patterns forecast future changes. The dynamic analysis can be applied to a natural resource-based industry or in fact to any industry that has shifted location over time and for which there is a direct link to a spatial dimension.

10.3 Final Remarks

The findings of this research are congruent with spatial predictions drawn by Location Theory, New Economic Geography and to some extent the Theory of Clusters. Resource-based industries are attracted to the location where their main input is plentiful and available at the lowest cost. Firms place the highest importance on the availability of logs and their prices when locating a new sawmill since wood is the most important input to the lumber manufacturing process. Thus, firms tend to locate closer to the source of inputs but at relative far distances from markets.

Because of this spatial arrangement access to roads to access markets and availability of transportation services are also important variables considered by decision-makers. Other important variables to decision makers in the softwood lumber industry are wages and energy costs as these are other necessary inputs to the manufacturing process. The conjoint analysis of site attribute preferences captured these conditions and the Ordered Probit and Logit models are a good approach to modeling location preferences. The most important attributes as rated in self-reported surveys are cost and availability of logs, followed by energy, labor and land costs.

Quality of roads is also important but to a lesser extent compared to the above mentioned attributes. The least important attribute of those included in the conjoint analysis, as expected, is distance to markets.

The spatial econometric model provided further evidence of these findings. The selection of the model was guided by a common factor analysis that identified six major factors influencing location in the softwood lumber sector. The variables that capture the effects of costs of logs and availability of logs showed to be statistically significant and had the expected signs when modeling the likelihood of softwood lumber occurrence in a county in the U.S. South. These findings reinforce the results of the conjoint analysis. The variable with the largest coefficient is the presence of a highway which is indicative of venues that allow access to markets. Access to roads is a major factor allowing for low-cost product transportation. Low transportation costs per unit have allowed the current spatial distribution of the softwood lumber industry as foreseen by Location Theory and New Economic Geography.

Cluster theory predicts the agglomeration of firms in a particular location because of the existence of competitive advantages. Other major factor facilitating the emergence of industrial clusters is access to centers of information and research. Although decision makers in the U.S. Pacific and Southern regions consider the industry to have cluster characteristics this does not meet all conditions specified in Porter (1998b)'s diamond model for competitive advantage. Access to labor is a necessary condition to the development of and industrial cluster and based on the result of the spatial econometric model, areas where labor is available coincides with the presence of the softwood lumber industry. However, firms do not locate near university campuses or final markets as denoted by the non-statistical significance of the proxies for these variables in the spatial econometric model. Proximity to a university for research support and availability of technical training for workers in their current location were rated among the least

important attributes for decision-makers. The softwood lumber industry still competes based on a least-cost strategy, and hence, the basis for competitive advantage on innovation for the development of a cluster does not seem to apply to this case.

The softwood lumber industry agglomerates in particular regions of the country mainly linked to the availability of input materials. The most important drivers of industry agglomeration are vertical integration, the availability of local suppliers and opportunities for improved innovation as rated by survey respondents. The strongest forces halting further clustering are increased competition and potential higher log prices. These forces resemble the centripetal and centrifugal forces of the New Economic Geography core-periphery model (Krugman 1995ab, Fujita and Krugman 2004). A common factor analysis of variables affecting industry agglomeration allows for the clear identification of such forces.

The strength of the spatial process in the econometric model is bleak despite reports in the literature that justify the use of a spatial structure due to inherent spatial processes and correlated errors because of the different scales of the information. It was determined that the strength of the spatial process declines as more variables with a spatial dimension are included. Although the theoretical appeal of a spatial autocorrelated model, empirical results show little differences in the results comparing a spatially correlated versus a conventional non-spatial econometric model. At the regional level spatial effects may be simply captured through the introduction of state specific and geographic coincidence variables than through a spatial scalar in a regression model.

Any new developments in sawmill capacity in the softwood lumber industry in the U.S. should take place in the U.S. South. There is evidence of industry spatial concentration in particular regions when comparing the distribution of firms for 1999 and 2005. The industry has

experienced considerable consolidation as demonstrated by a reduction in the number of firms accompanied with an increase in sawmill capacity in the U.S. South and Texas.

This multi-disciplinary framework used approaches and tools borrowed from the marketing, econometrics, GIS, and spatial statistics literature. It constitutes a new approach to examine industry location decisions and behavior. Research methods have the ability to capture decision-makers preferences and to provide evidence in favor of the major theories involving location, economic geography and cluster development. Results can provide industry and economic development professionals with a new decision-making tool that can help identify areas where new industry development could occur in the future with a high probability of success.

LITERATURE CITED

Adamowicz, W., Louviere, J., Swait, J. (1998). Introduction to Attribute-Based Stated Choice Methods. National Oceanic and Atmospheric Administration. 44 p.

Aguilar, F.X. and R.P. Vlosky. (2006). Spatial Analysis of Forest Product Manufacturer Clusters in Louisiana. Wood and Fiber Science. 38(1): 121-131.

Albert, J. H. and S. Chib (1993). Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association 88(422): 669-679.

Alig, R. J., D. M. Adams, et al. (1998). Ecological and economic impacts of forest policies: interactions across forestry and agriculture. Ecological Economics 27(1): 63-78.

Alig, R.J., Hohenstein, W.G., Murray, B.C., and R.G. Haight. Changes in Area of Timberland in the United States, 1952-2040, By Ownership, Forest Type, Region and State. United States Department of Agriculture. Forest Service. Southeastern Forest Experiment Station. General Technical Report SE-64. 34p .

Luc Anselin, Ibnu Syabri, and Youngihn Kho (2005). GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38 (1): 5–22.

Anselin, L. (2000). Part 2 The Link between GIS and spatial analysis: GIS, spatial econometrics and social science research. Journal of Geographical Systems (5):11-15.

Anselin, L. (2001). Spatial Effects in Econometric Practice in Environmental and Resource Economics. American Journal of Agricultural Economics. 83(3): 705-710.

Anselin, L. and Griffith D.A. (1988). Spatial Econometrics: Methods and Models. Kluwert, Dordrecht. 284p.

Anselin, L. and A. Getis. (1992). Spatial statistical analysis and geographic information systems. Annals of Regional Science 26(1): 19-33.

Anselin, L. and S. Hudak. (1992). Spatial econometrics in practice: A review of software options. Regional Science and Urban Economics 22(3): 509-536.

Armstrong, J.S. and T.S. Overton. (1977). Estimating non-response bias in mail surveys. Journal of Marketing Research. 14(3): 396-402.

Arnold, S.J., Roth, V., Tigert, D.J. (1981). Conditional Logit versus MDA in the Prediction of Store Choice. Advances in Consumer Research. 8(1):665-670.

Austin, J. and D. Lozano. (1975). The Jones Act and the Douglas-fir Region Softwood Lumber Industry in Perspective. Journal of Forestry. 73(10): 644-648.

Austin, J. and G., Lozano. (1999). Mexican Foundation for Rural Development. Harvard Business School Cases.14 p.

Baker, O. E., R. T. Ely, et al. (1926). Economics and Geography. The American Economic Review 16(1): 112-115.

Barbe, G. (1993). Methods of Transporting Timber in the Southern United States. Louisiana Forest Products Development Center. Internal document. 23 p.

Baumol, W. J. (1968). Entrepreneurship in Economic Theory. The American Economic Review 58(2, Papers and Proceedings of the Eightieth Annual Meeting of the American Economic Association): 64-71.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. 36(2): 192-236.

Bethel, J.S. and G.F. Schreuder. (1976). Forest Resources: An Overview. Science Vol. 191: 747-752.

Bigsby, H.R. (1994). Production structure and the Australian Sawmilling industry. Australian Journal of Agricultural Economics. 38(3):271-288.

Blackman, A., Avalos-Sartorio, B., Chow, J. and F.X. Aguilar. 2006. Tree cover loss in El Salvador's coffee forests. Report to the Inter-American Development Bank. RFF, Washington D.C. 97 p. Online at http://www.rff.org/Documents/RFF-Rpt-TreeCoverLoss.pdf. Accessed December 15, 2006.

Braden, R., H. Fossum, I.Eastin, J. Dirks, and E. Lowell. (1998). The role of manufacturing clusters in the Pacific Northwest forest products industry. Center for International Trade in Forest Products. Seattle, WA. Working paper 66. 43 p.

Bretton-Clark. Conjoint Designer. New York: Bretton-Clark Co., 1988.

Bruner, G.C., James, K.E., Hensel, P.J. (2001). Marketing Scales Handbook : A compilation of multi-item measures. Volumen III. Chicago, Ill., USA , American Marketing Association. 1650 p.

Carroll, M. S., C. W. McKetta, Blatnet, K.A., and C. Schallau. (1999). A Response to "Forty Years of Spotted Owls? A Longitudinal Analysis of Logging Industry Job Losses. Sociological Perspectives 42(2): 325-333.

Chance, N.P and Vlosky, R.P. (1995). Louisiana's Forests and Forest Products Industries. Louisiana Forest Products Laboratory. Working Paper #11. Louisiana State University Agricultural Center. 10 p.

Chisholm, M. (1968). Johann Heinrich von Thünen. In: Readings in Economic Geography: The location of Economic Activity. R.H.T. Smith, E.J. Taaffe and L.J. King (Eds.). 34-40.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. Econometrica 28(3): 591-605.

Ciccone, A. and R. E. Hall (1996). Productivity and the Density of Economic Activity. The American Economic Review 86(1): 54-70.

Cliff, A.D. and J.K. Ord. (1981). Spatial Processes Models and Applications. Pion, London.

Cohen, J. P. and C. J. M. Paul (2005). Agglomeration economies and industry location decisions: the impacts of spatial and industrial spillovers. Regional Science and Urban Economics 35(3): 215-237.

Cressie, N.A.C. (1993). Statistics for Spatial Data. Wiley, NY. 900 p.

Diamond, C. A. and C. J. Simon (1990). Industrial Specialization and the Returns to Labor. Journal of Labor Economics 8(2): 175-201.

Dillman, D.A. (1978). Mail and Telephone Surveys: The Total Design. John Wiley and Sons, New York.

Dillman, D.A. (2000). Mail and Internet Surveys: The Total Design Method. John Wiley and Sons, New York.

Dirks, J.M., and D.G. Briggs. (1991). Wood Products in Washington State: The Secondary Manufacturing Industries. Center for International Trade in Forest Products.Working Paper 30. University of Washington, Seattle WA.

Dorfman, J.H., 1996. Modeling Multiple Adoption Decisions in a Joint Framework. American Journal of Agricultural Economics 78, 547-557.

Ellison, G. and E. L. Glaeser (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. Journal of Political Economy 105(5): 889-927.

Elrod, T., Louveire, J.L., Davey, K. (1992). An empirical comparison of ratings-based and choice-based conjoint models. Journal of Marketing Research Vol. 29: 368-377.

Fingleton, B. (2001). Theoretical economic geography and spatial econometrics: dynamic perspectives. Journal of Economic Geography.1(2): 201-225.

Fischer, M. M. and P. Nijkamp (1992). Geographic information systems and spatial analysis. Annals of Regional Science 26(1): 3-17.

Fleming, M.M. (2004) Techniques for Estimating Spatially Dependent Discrete Choice Models. In Advances in Spatial Econometric. Methodology, Tools and Applications. Edited by L. Anselin, RJ.G.M. Florax and S.J. Rey Springer. 145-168

Florax, R. and T. de Graaf. (2004). The Performance of Diagnostic Tests for Spatial Dependence in Linear Regression Models: A Meta-Analysis of Simulation Studies. In Advances in Spatial Econometric. Methodology, Tools and Applications. Edited by L. Anselin, RJ.G.M. Florax and S.J. Rey Springer. 29-66.

Florax, R. and H. Folmer (1992). Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators. Regional Science and Urban Economics 22(3): 405-432.

Fujita, M. and P. Krugman (2004). The new economic geography: Past, present and the future. Papers in Regional Science 83(1): 139-164.

Gensch, D. H. and W. W. Recker (1979). The Multinomial, Multiattribute Logit Choice Model. Journal of Marketing Research. 16(1): 124-132.

Getis, A. (2000). Introduction: Spatial analysis and GIS: An introduction. Journal of Geographical Systems 2(1): 1-3.

Goodchild, M. F. (2000). Part 1 Spatial analysts and GIS practitioners: The current status of GIS and spatial analysis. Journal of Geographical Systems: (2):5-10.

Goodchild, M. F. and R. P. Haining. (2004). GIS and spatial data analysis: Converging perspectives. Papers in Regional Science 83(1): 363-385.

Granskog, J.E., Haines, T., Greene, J.L., Doherty, B.A., Bick, S., Haney Jr., H.L., Moffat, S.O., Speir, J. and J. J. Spink. Chapter 8: Policies, Regulations, and Laws. In: Wear, David N.; Greis, John G., eds. 2002. Southern forest resource assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. p.189-223.

Green, P.E., and V. Srinivasan. (1978). Conjoint Analysis in Consumer Research: Issues and Outlook. Journal of Consumer Research. Vol. 5:103-113.

Green, P.E., and V. Srinivasan. (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. Journal of Marketing. Vol. 4:3-19.

Green, P.E. and Y.Wind. (1975). New way to measure consumer' judgments. Harvard Business Review. July-August. 53(4): 107-117.

Greene, W.H. (2003). Econometric Analysis. 5th edition. Pearson Education, 1026 p.

Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation problem. Journal of Geographical Systems 2(2): 141-156.

Hamilton, F. E. I. (1971). Decision-Making and Industrial Location in Eastern Europe. Transactions of the Institute of British Geographers(52): 77.

Haining, R. (1990). Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press. 409 p.

Haining, R. (2003). Spatial Data Analysis. Theory and Practice. Cambridge University Press. 432 p.

Hair, J.F., Anderson, R.E., Tathan, R.L. and W.C. Black. (1998). Multivariate Data Analysis. 3rd ed. MacMillan Publishing Co. New York. 730 pp.

Halbrendt, C.K., Wirth, F.F., and G.F. Vaughn. 1991. Conjoint Analysis of the Mid-Atlantic Food-Fish Market for Farm Raised Hybrid Striped Bass. Southern Journal of Agricultural Economics. Vol. 23: 155-163.

Hallencreutz, D. and P. Lundequist. (2003). Spatial clustering and the potential for policy practice: experiences from cluster-building processes in Sweden. European Planning Studies. 11(5):533-548.

Halley, E. (1686). An Historical Account of the Trade Winds, and Monsoons, Observable in the Seas between and Near the Tropics, with an Attempt to Assign the Physical Cause of the Said Winds. Philosophical Transactions (1683-1775) 16: 153-168.

Hanemann, W.M. (1984). Discrete/Continuous Models of Consumer Demand. Econometrica. 52(3): 541-561.

Hansen, F. (1969). Consumer Choice Behavior: An Experimental Approach. Journal of Marketing Research 6(4): 436.

Harris, T.G. (1988) Wood Procurement Systems Overview. In: Current Challenges to Traditional Wood Procurement Systems. Ed. by M.P. Hamel. Forest Products Research Society Proceedings 47353. Atlanta, Georgia, September 20-30, 1987. p. 27-29.

Harrison, R.W., Stringer, T., and Prinyawiwatkul, W. (2002). An analysis of consumer preferences for value-added seafood products derived from crawfish. Agricultural and Resource Economics Review. Vol 31: 157-170.

Harrison, R. W., Gillespie, J., and D. Fields. (2005). Analysis of cardinal and ordinal assumptions in conjoint analysis. Agricultural and resource economics review. Vol 34: 238-252.

Harrison, R.W., and E. Mclennon. (2004). Analysis of Consumer Preferences for Biotech Labeling Formats. Journal of Agricultural and Applied Economics 36(1): 159-171.

Harrison, R.W., and P. Sambidi. (2004). A Conjoint Analysis of the U.S. Broiler Complex Location Decision. Journal of Agricultural & Applied Economics. 36(3): 639-655.

Hausman, J. and D. McFadden (1984). Specification Tests for the Multinomial Logit Model. Econometrica 52(5): 1219-1240.

Helburn, S. (1943). Location of Industry. The Journal of Land & Public Utility Economics 19(3): 253-263.

Henig, J. R. and J. A. MacDonald (2002). Locational Decisions of Charter Schools: Probing the Market Metaphor. Social Science Quarterly. 83(4): 962-980.

Hoover, E. (1948). The Location of Economic Activity. McGraw-Hill, New York.310 p.

Hotelling, H. (1929). Stability in Competition. The Economic Journal. 39(153): 41-57.

Hovee E.D and Logue, A. (2005). Oregon Forest Cluster Analysis. Report prepared for: Oregon Forest Resources Institute

Isard, W. (1949). The General Theory of Location and Space-Economy. The Quarterly Journal of Economics 63(4): 476-506.

Jaffe, A.B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. American Economic Review, American Economic Association. 76(5): 984-1001.

Jaffe, A. B. (1989). Real Effects of Academic Research. American Economic Review. 79(5): 957-970.

Jaffe, A. B. and M. Trajtenberg (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108(3): 577-598

Jaffe, A. B., M. Trajtenberg, et al. (2000). Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. American Economic Review 90(2, Papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American Economic Association): 215-218.

Johnson, R.A. and D.W. Wichern. (2002). Applied Multivariate Statistical Analysis. 5th edition. Prentice Hall: Upper Saddle River, New Jersey. 767 p.

Kaluzny, S.P., S. Vega, T. Cardoso, A. Shelly. (1997). S+Spatial Stats: Users' manual for Windows and Unix. Springer, Seattle, WA. 327 p.

Keasbey, L. M. (1901). The Study of Economic Geography. Political Science Quarterly 16(1): 79-95.

Keegan III, C., Morgan, T.A., Gebert, K.M., Brandt, J.P., Blatner, K.A. and T.P. Spoelma. (2006). Timber-processing Capacity and Capabilities in the Western United States. Journal of Forestry. 104 (5): 262-268.

Kennelly, R.A. (1968). The Location of the Mexican Steel Industry. In: Readings in Economic Geography: The location of Economic Activity. R.H.T. Smith, E.J. Taaffe and L.J. King (Eds.). 126-157.

Kilkenny, M. (1998). Transport Costs, the New Economic Geography, and Rural Development. Growth and Change. Vol. 29: 259-280.

Krugman , P. (1991). Increasing Returns and Economic Geography. Journal of Political Economy. 99(3):483-499.

Krugman, P. (1995a). Development, Geography, and Economic Theory. Chapter 2: Geography Lost and Found. Cambridge, MIT Press. p.31-65.

Krugman, P. (1995b). Development, Geography and Economic Theory. Chapter 5: Core and periphery. MIT Press. p.61-77.

Krugman, P. (1998). The Role of Geography in Development. Annual World Bank Conference on Development Economics. Washington, D.C., April 20–21, 1998. 34 p.

Kuhfeld, W. F.; Tobias, R.D., Garratt, M. (1994). Efficient experimental design with marketing research applications. Journal of Marketing Research. 31(4):545-547.

Kuhfeld, W.F. (1993). SAS Technical Report R-109. Conjoint Analysis Examples. SAS Institute Inc. 85 pp.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. Journal of Political Economy. 74(2): 132-157.

Lee Abt, K., Winter, S.A. and R. J. Huggett, Jr. Chapter 10: Local Economic Impacts of Forests. In: Wear, David N.; Greis, John G., eds. 2002. Southern forest resource assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. p 239-267.

LeSage, J. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. Geographical Analysis. 32: 19-35.

Leung, Y. (2000). Some thoughts on spatial data and structure analysis. Journal of Geographical Systems. 2(1): 107-110.

Long, J.S. and J. Freese. (2006). Regression Models for Categorical Dependent Variables Using Stata. State Press. 527 p.

Louviere, J.J., Hensher, D.A., Swait, J.D. (2000). Stated Choice Methods. Analysis and Application. Cambridge University Press. 402 p.

Lusk, J.L., Norwood, F.B. (2005). Effect of Experimental Design on Choice-Based Conjoint Valuation Estimates. American Journal of Agricultural Economics. 87(3):771-785.

Mackenzie, J. (1993). A comparison of contingent preference models. American Journal of Agricultural Economics. 75(3): 593-603.

Marshall. A. (1920). Principles of Economics. Macmillan, London. 820 p.

McCann, P. and S. Sheppard (2003). The Rise, Fall and Rise Again of Industrial Location Theory. Regional Studies 37(6/7): 649-663.

McCauley, C.K. and J.P. Caulfield. (1990). Using mixed-integer programming to determine the optimal location for an oriented strandboard plant in Alabama. Forest Products Journal. 10(2):39-44.

McFadden, D.L. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In Frontier in Econometrics. Edited by P. Zarembka. New York, Academic Press, 105-142.

McFadden, D. (1986). The Choice Theory Approach to Market Research. Marketing Science. 5(4):275-297.

McMillen, D. P. (1992). Probit With Spatial Autocorrelation. Journal of Regional Science 32(3): 335-348.

McMillen, D. P. (2003). Spatial Autocorrelation or Model Misspecification. International Regional Science Review 26(2): 208-217.

McRAE's Blue Book. (2006). Available at http://www.macraesbluebook.com. Accessed October 5, 2006.

Munroe, D. K., J. Southworth, et al. (2004). Modeling Spatially and Temporally Complex Land-Cover Change: The Case of Western Honduras. Professional Geographer 56(4): 544-559.

Murray, B. (1995). Measuring oligopsony power with shadow prices: U.S. markets for pulpwood and sawlogs. Review of Economics & Statistics. 77:486-499.

Murray, B. C. and D. N. Wear (1998). Federal Timber Restrictions and Interregional Arbitrage in U.S. Lumber. Land Economics 74(1): 76-91.

NationalAtlas.Gov. (2004). Climate: Average Annual Precipitation, 1961-1990. Available at http://www.nationalatlas.gov/atlasftp.html. Accessed July 30, 2006.

Nelson, G. C. and D. Hellerstein (1997). Do roads cause deforestation? Using satellite images. American Journal of Agricultural Economics. 79(1): 80-88.

Peneder, M. (1995). Cluster techniques as a method to analyze industrial competitiveness. International Advances in Economic Research. 1(3): 295-304.

Pinkse, J. and M. E. Slade (1998). Contracting in space: An application of spatial statistics to discrete-choice models. Journal of Econometrics 85(1): 125-154.

Porter, M. E. (1998a). The Adam Smith address: Location, clusters, and the `new' Microeconomics of Competition. Business Economics 33(1): 7-13.

Porter, M.E. (1998b). Clusters and Competition: New Agendas for Companies, Governments, and Institutions. Harvard Business School Press. Product # 2034. 54 p.

Porter, M. E. (1998c). Clusters and the New Economics of Competition. Harvard Business Review 76(6): 77-90.

Porter, M.E. (2000). Location, Competition and Economic Development: Local Clusters in a Global Economy. Economic Development Quarterly. 14(1):15-34.

Porter, M. E. (2003). The Economic Performance of Regions. Regional Studies 37(6/7): 545.

Predohl, A. (1928). The Theory of Location in Its Relation to General Economics. The Journal of Political Economy 36(3): 371-390.

Prestemon, J.P. and Abt, R.C. Chapter 13: Timber Products Supply and Demand In: Wear, David N.; Greis, John G., eds. 2002. Southern forest resource assessment. Gen. Tech. Rep. SRS-53. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. p 299-325.

Punj, G. N. and R. Staelin. (1978). The Choice Process for Graduate Business Schools. Journal of Marketing Research: 11:588-598.

Raftery, A.E. 1996. Bayesian model selection in social research. In Sociological Methodology. Edited by P.V. Marsden. VOl. 26: 111-163.

Random Lengths. (2006). Big Book 2005.

Rawstron, E. M. (1958). Three Principles of Industrial Location. Transactions and Papers (Institute of British Geographers)(25): 135-142.

Renner, G. T. (1931). Geographic Regions of Sierra Leone. Economic Geography 7(1): 41-49.

Ringe, J. M. and W.L. Hoover. (1987). Value added analysis: a method of technological assessment in the U.S. forest products industry. Forest Products Journal. 37(11/12):51-54.

Robbins, W. G. (1985). The Social Context of Forestry: The Pacific Northwest in the Twentieth Century. The Western Historical Quarterly. 16(4): 413-427.

Roe, B. and K. J. Boyle (1996). Using conjoint analysis to derive estimates of compensating variation. Journal of Environmental Economics & Management 31(2): 145-159.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. The Journal of Political Economy 82(1): 34-55.

Ross, E. A. (1896). The Location of Industries. The Quarterly Journal of Economics 10(3): 247-268.

Rummel, R.J. 1970. Applied Factor Analysis. Northwestern University Press. 617 p.

Seldon, B.J and S. Bullard. (1992). Input substitution, economies of scale and productivity growth in the US upholstered furniture industry. Applied Economics. Vol. 24: 1017-1025.

Smith, M. D. (2002). Two Econometric Approaches for Predicting the Spatial Behavior of Renewable Resource Harvesters. Land Economics 78(4): 522-538.

Society of American Foresters. (2006). 2006 SAF-Accredited Professional Forestry Degree Programs. Online at http://www.safnet.org/education/pforschools.cfm. Accessed October 10, 2006.

Society of Wood Science and Technology. (2006). SWST Directory North American Schools offering Baccalaureate and Graduate Programs of Study in Wood Science and Technology. Online at http://www.swst.org/schooldirectory.html. Accessed October 10, 2006.

Spelter, H. (2003). Sawmill closures, openings, and net capacity changes in the softwood lumber sector, 1996-2003. USDA Forest Products Laboratory, Madison, Wisconsin. 14 pp. Online at http://www.fpl.fs.fed.us/documnts/fplrp/fplrp603.pdf. Accessed October 10, 2006.

Spelter, H., and Alderman, M. (2005). Profile 2005: Softwood sawmills in the United States and Canada. Research Paper FPL-RP-630. U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 85 pp. Online at http://www.fpl.fs.fed.us/documnts/fplrp/fpl_rp630.pdf. Accessed October 24, 2006.

Spelter, H. and M. Alderman. (2003) Profile 2003: Softwood Sawmills in the United States and Canada. Research Paper FPL-RP-608. U.S. Department of Agriculture, Forest Service. Forest Products Laboratory. 79 pp. Online at http://www.fpl.fs.fed.us/documnts/fplrp/fplrp608.pdf. Accessed October 24, 2006.

Stokes, B.J. (1997). Southern regional report. In: Forest operations for sustainable forests and healthy economies; Proceedings, Council on Forest Engineering, 20th annual meeting; 1997 July 28-31; Rapid City, SD. Rapid City, SD: Council on Forest Engineering: 172-174.

Student. (1907). On the Error of Counting with a Haemacytometer. Biometrika 5(3): 351-360.

Swait, J., and W. Adamowicz. (2001). The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. Journal of Consumer Research. Vol. 28 (1):135-148.

Sy, H. A. and M. D. Faminow (1997). Estimating the values of cattle characteristics using an ordered probit model. American Journal of Agricultural Economics 79(2): 463-476.

Theil, H. (1970). On the Estimation of Relationships Involving Qualitative Variables. The American Journal of Sociology 76(1): 103.

Thompson, P. and M. Fox-Kean (2005). Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. American Economic Review 95(1): 450-460.

Thurstone, L. L. (1927). Psychophysical Analysis. The American Journal of Psychology 38(3): 368-389.

Timber Mart-South.(2006). Delivered Prices. The Journal of Southern Timber Prices .1st Quarter 2006. Vol. 31 No.1.

Timber Mart-South.(2006). Delivered Prices. The Journal of Southern Timber Prices. 2nd Quarter 2006. Vol. 31 No.2.

U.S. Census Bureau. (2005). Geographic Area Statistics: 2003. Annual Survey of Manufacturers. U.S. Department of Commerce. Economics and Statistics Administration. 110 p. Available at http://www.census.gov/prod/2005pubs/am0331as1.pdf. Accessed August 7, 2006.

U.S. Census Bureau. (2006a). MA321T: Lumber Production and Mill Stocks: 2005. 13 pp. Online at http://www.census.gov/industry/1/ma321t05.pdf. Accessed October 10, 2006.

U.S. Census Bureau. (2006b). Statistics for Industry Groups and Industries: 2005. Annual Survey of Manufactures M05(AS)-1. 340 pp. Available at http://www.census.gov/prod/2006pubs/am0531gs1.pdf. Accessed October 10, 2006.

U.S. Department of Transportation (1998). North American Transportation Atlas Data. Available at http://www.bts.gov/publications/north_american_transportation_atlas_data/. Accessed August 18, 2006.

U.S. Department of Agriculture National Agricultural Statistics Service. (2004). Volume 1 Geographic Area Series. Census, State - County Data. Available at http://www.nass.usda.gov/Census_of_Agriculture/index.asp. Accessed September 15, 2006.

U.S. Department of Agriculture, Forest Service Southern Research Station. 2005. U.S. Wood-Using Mill Locations, SRS-4850. Available at http://www.srs.fs.usda.gov/econ/data/mills/mill2005.htm#downloads. Accessed October 15, 2006.

Vaughn, G. F. (1994). The Geography of Resource Economics. Land Economics. 70(4): 515-519.

Vlosky, R.P. (1995). An Overview of the Louisiana Primary Solid Wood Products Industry. Louisiana Forest Products Laboratory. Working Paper # 2. Louisiana State University Agricultural Center. 18 p.

Vlosky, R.P., Chance, N.P., and O.V. Harding. (1994). An Overview of the Louisiana Secondary Solid Wood Products Industry. Louisiana Forest Products Laboratory. Working Paper # 1. Louisiana State University Agricultural Center. 33 p.

Vlosky, R. P. and N. P. Chance (2001). Employment Structure And Training Needs In The Louisiana Value-Added Wood Products Industry. Forest Products Journal. 51(3):34-41.

Vlosky, R. P. and L. K. Ozanne (1998). Environmental certification of wood products: The U.S. manufacturers' perspective. Forest Products Journal 48(9): 21-26.

Vlosky, R. P. and T. F. Shupe (2004). Buyer perceptions and purchasing patterns related to treated wood use in children's playground equipment. Forest Products Journal 54(12): 307-312.

Vlosky, R. P. and T. F. Shupe (2004). An exploratory study of home builder, new-home homeowner, and real estate agent perceptions and attitudes about mold. Forest Products Journal 54(12): 289-295.

Vlosky, R. P., T. Westbrook and K. Poku. (2002). An Exploratory Study of Internet Adoption by Primary Wood Products Manufacturers in the Western United States. Forest Products Journal. 52(6):35-42.

Vlosky, R. P. and Q. Wu. (2001).A Brief Look at Raw Material Usage in the Furniture and Cabinet Industries in the Southern United States. Forest Products Journal. 51(9)25-29.

Warren, D. D. (2003). Production, prices, employment, and trade in Northwest forest industries, all quarters of 2001. Resources Bulletin. PNW-RB-239. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 171 p. Online at http://www.fs.fed.us/pnw/pubs/rb239/rb239.pdf. Accessed October 25, 2006.

Warren, D. D. (1990). Production, prices, employment, and trade in Northwest forest industries, first quarter 1990. Resources Bulletin. PNW-RB-175. Portland, OR: US. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 99 p Online at http://www.fs.fed.us/pnw/pubs/pnw_rb175.pdf. Accessed October 25, 2006.

Warren, D. D. (1987). Production, prices, employment, and trade In Northwest forest industries, Fourth quarter 1986. Resources Bulleting. PNW-RB-144. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station; 1987. 55 p. Online at http://www.fs.fed.us/pnw/pubs/pnw_rb144.pdf. Accessed October 25, 2006.

Wear, D.N. and B. C. Murray (2004). Federal timber restrictions, interregional spillovers, and the impact on US softwood markets. Journal of Environmental Economics and Management 47(2): 307.

Wear, D.N., Carter D.R. and J. Prestemon. (2005). The US South’s Timber Sector in 2005: A prospective analysis of recent change. USDA, Forest Service. Available at

http://www.srs.fs.usda.gov/econ/pubs/southernmarkets/southern-markets-01.htm. Accessed August 7, 2006.

Weber, A. (1929).Theory of the Location of Industries. The University of Chicago Press. 256 p.

Whittle, P. (1954). On Stationary Processes in the Plane. Biometrika 41(3/4): 434-449.

Wilkie, W. L. and E. A. Pessemier (1973). Issues in Marketing's Use of Multi-Attribute Attitude Models. Journal of Marketing Research 10(4): 428.

Williamson, O. E. (1963). Managerial Discretion and Business Behavior. The American Economic Review 53(5): 1032-1057.

Wooldridge, J.M.. (2002). Econometric Analysis of Cross Section and Panel Data. The MIT Press. 752 p.

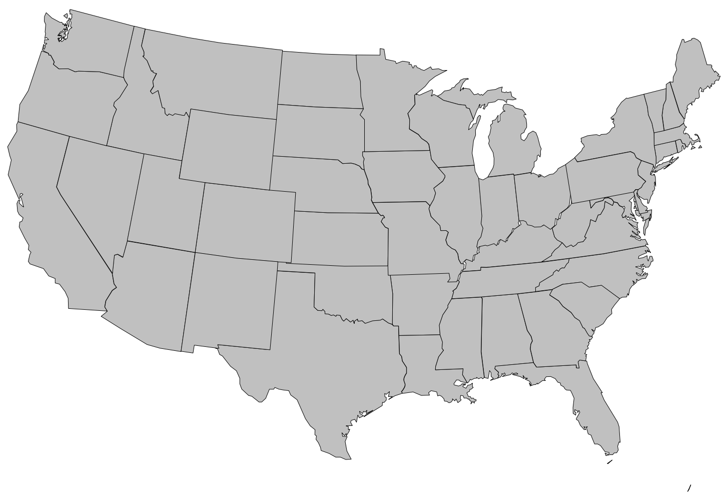
Wooldridge, J. M. (2003). Cluster-Sample Methods in Applied Econometrics. The American Economic Review. 93(2): 133-138.

APPENDIX: SURVEY INSTRUMENT

Making Decisions on Locating Softwood Sawmills in the U.S.

Louisiana Forest Products Development Center

LSU AgCenter Research & Extension



Please take 10 minutes to complete this short survey and
return it to us by November 5th in the enclosed postage paid envelope

If you have any questions about this survey, please contact Francisco X. Aguilar, Doctoral Candidate, Forest Products Marketing, Louisiana Forest Products Development Center, School of Renewable Natural Resources. Louisiana State University Agricultural Center, Baton Rouge, LA 70803; Phone: (225) 578-4133; Fax (225) 578-4251; e-mail: faguil1@lsu.edu

October 2006

Section I. Background Information-YOUR MILL

1. Is softwood lumber manufactured at this location? (please check one)

☐ **YES** ☐ **NO →** **If NO, please *stop* and return in the postage paid envelope**

---

2. Please indicate the *physical* address for this mill.

Address: ____

City: _________________ State: _________ Zip code: _______________

3. Please indicate the total number of <u>full-time</u> employees at this mill. (Please check only one response)

☐ **5 or less**	☐ **50-74**
☐ **6-10**	☐ **75-99**
☐ **10-24**	☐ **100-149**
☐ **25-49**	☐ **150 or more**

4. Please estimate 2005 sales revenues from THIS MILL. (Please check only one response)

☐ **Less than $10 million**	☐ **$60 – $69.9 million**
☐ **$10 – $19.9 million**	☐ **$70 – $79.9 million**
☐ **$20 – $29.9 million**	☐ **$80 – $89.9 million**
☐ **$30 – $39.9 million**	☐ **$90 – $99.9 million**
☐ **$40 – $49.9 million**	☐ **$100 – $109.9 million**
☐ **$50 – $59.9 million**	☐ **$110 million OR MORE**

5. Please estimate 2005 annual production (in Million Board Feet) from THIS MILL.

__________ MMBF

6. How many years has been this sawmill in operation?

___________ years

7. Please indicate which best describes your position(s) in your company. (Please check all that apply)

☐ **Owner**	☐ **Marketing Manager**
☐ **Sales Manager**	☐ **Plant Manager**
☐ **Other (please specify)**______________________	

Section II. Softwood Sawmill Site Location Factors-Your Mill

1. **Please indicate the importance of the following factors that influence where *this* mill is located. (Circle only one for each).**

	Not Important at All	Somewhat Unimportant	Neither Unimportant Nor Important	Somewhat Important	Very Important
Cost of land	1	2	3	4	5
Cost of logs	1	2	3	4	5
Sufficient supply of logs	1	2	3	4	5
Cost of energy	1	2	3	4	5
Sufficient supply of energy	1	2	3	4	5
Regional average wages	1	2	3	4	5
Non-skilled labor availability	1	2	3	4	5
Skilled labor availability	1	2	3	4	5
Quality of roads	1	2	3	4	5
Rail and railcar availability	1	2	3	4	5
Proximity to ports	1	2	3	4	5
Distance to markets	1	2	3	4	5
Proximity to log supply area	1	2	3	4	5
Trucks and trucking availability	1	2	3	4	5
Lack of competition from other sawmills	1	2	3	4	5
Proximity to a university for research support	1	2	3	4	5
Availability of technical training for workers	1	2	3	4	5
Availability of State business incentives	1	2	3	4	5
Favorable environmental regulations	1	2	3	4	5
Favorable state property taxes	1	2	3	4	5
Favorable local property taxes	1	2	3	4	5
Favorable State fuel taxes	1	2	3	4	5
Proximity to health care services	1	2	3	4	5
Quality of education for worker's families	1	2	3	4	5

Please list any other criteria you think are important for your mill's location:

Section III. NEW Softwood Sawmill Site Location Decisions

*In this section we would like to know your preferences for different characteristics if you had to decide where to locate a **NEW** sawmill.*

In each of the boxes below and on page 5, Profiles A and B represent hypothetical characteristics for the location of a **NEW SOFTWOOD SAWMILL**. Please select the profile that you prefer in each box.

Please select A or B:

Characteristics	Profile: A	Profile: B
Average hourly wage in the region	$10.50/hour	$10.50/hour
Average price for logs	$62.25/ton	$46.31/ton
Electricity cost	4.50 cents/kWh	6.50 cents/kWh
Cost of land where mill is to be situated	Low	High
Quality of roads	Poor	Poor
Distance to source for logs	70 miles	30 miles
Distance to market	90 miles	90 miles
Please check your preferred choice →	☐ A	☐ B

Please select A or B:

Characteristics	Profile: A	Profile: B
Average hourly wage in the region	$15.50/hour	$15.50/hour
Average price for logs	$62.25/ton	$46.31/ton
Electricity cost	6.50 cents/kWh	4.50 cents/kWh
Cost of land where mill is to be situated	High	High
Quality of roads	Good	Poor
Distance to source for logs	70 miles	70 miles
Distance to market	90 miles	20 miles
Please check your preferred choice →	☐ A	☐ B

Please continue with boxes
on the next page

→

Please select A or B:

Characteristics	**Profile: A**	**Profile: B**
Average hourly wage in the region	**$15.50/hour**	**$10.50/hour**
Average price for logs	**$62.25/ton**	**$62.25/ton**
Electricity cost	**6.50 cents/kWh**	**4.50 cents/kWh**
Cost of land where mill is to be situated	**Low**	**High**
Quality of roads	**Poor**	**Good**
Distance to source for logs	**30 miles**	**30 miles**
Distance to market	**20 miles**	**20 miles**
Please check your preferred choice →	☐ **A**	☐ **B**

Please select A or B:

Characteristics	**Profile: A**	**Profile: B**
Average hourly wage in the region	**$15.50/hour**	**$10.50/hour**
Average price for logs	**$46.31/ton**	**$46.31/ton**
Electricity cost	**4.50 cents/kWh**	**6.50 cents/kWh**
Cost of land where mill is to be situated	**Low**	**Low**
Quality of roads	**Good**	**Good**
Distance to source for logs	**30 miles**	**70 miles**
Distance to market	**90 miles**	**20 miles**
Please check your preferred choice →	☐ **A**	☐ **B**

Please review the 8 boxes shown below and on Page 7. Each box contains a combination of location characteristics for a hypothetical **NEW SOFTWOOD LUMBER SAWMILL**. Please **RATE** each box with an "X" in the scale below it where 1 = *Not attractive at all*, and 7 = *Very attractive*.

Location characteristics
Average hourly wage in the region: $15.50/hour
Average price for logs: $62.25/ton
Electricity cost: 6.50 cents/kWh
Cost of land where mill is to be situated: High
Quality of roads from mill to main market: Good
Distance to source for logs: 70 miles
Distance to market: 90 miles

Not attractive at all :____:____:____:____ :____:____:____ Very attractive
1 2 3 4 5 6 7

Location characteristics
Average hourly wage in the region: $15.50/hour
Average price for logs: $46.31/ton
Electricity cost: 4.50 cents/kWh
Cost of land where mill is to be situated: Low
Quality of roads from mill to main market: Good
Distance to source for logs: 30 miles
Distance to market: 90 miles

Not attractive at all :____:____:____:____ :____:____:____ Very attractive
1 2 3 4 5 6 7

Location characteristics
Average hourly wage in the region: $15.50/hour
Average price for logs: $62.25/ton
Electricity cost: 6.50 cents/kWh
Cost of land where mill is to be situated: Low
Quality of roads from mill to main market: Poor
Distance to source for logs: 30 miles
Distance to market: 20 miles

Not attractive at all :____:____:____:____ :____:____:____ Very attractive
1 2 3 4 5 6 7

Location characteristics
Average hourly wage in the region: $15.50/hour
Average price for logs: $46.31/ton
Electricity cost: 4.50 cents/kWh
Cost of land where mill is to be situated: High
Quality of roads from mill to main market: Poor
Distance to source for logs: 70 miles
Distance to market: 20 miles

Not attractive at all :____:____:____:____ :____:____:____ Very attractive
1 2 3 4 5 6 7

Please continue with options on the next page

Location characteristics
Average hourly wage in the region: $10.50/hour
Average price for logs: $62.25/ton
Electricity cost: 4.50 cents/kWh
Cost of land where mill is to be situated: High
Quality of roads from mill to main market: Good
Distance to source for logs: 30 miles
Distance to market: 20 miles

Not attractive at all :_____:_____:_____:_____ :_____:_____:_____ Very attractive
1 2 3 4 5 6 7

Location characteristics
Average hourly wage in the region: $10.50/hour
Average price for logs: $46.31/ton
Electricity cost: 6.50 cents/kWh
Cost of land where mill is to be situated: Low
Quality of roads from mill to main market: Good
Distance to source for logs: 70 miles
Distance to market: 20 miles

Not attractive at all :_____:_____:_____:_____ :_____:_____:_____ Very attractive
1 2 3 4 5 6 7

Location characteristics
Average hourly wage in the region: $10.50/hour
Average price for logs: $62.25/ton
Electricity cost: 4.50 cents/kWh
Cost of land where mill is to be situated: Low
Quality of roads from mill to main market: Poor
Distance to source for logs: 70 miles
Distance to market: 90 miles

Not attractive at all :_____:_____:_____:_____ :_____:_____:_____ Very attractive
1 2 3 4 5 6 7

Location characteristics
Average hourly wage in the region: $10.50/hour
Average price for logs: $46.31/ton
Electricity cost: 6.50 cents/kWh
Cost of land where mill is to be situated: High
Quality of roads from mill to main market: Poor
Distance to source for logs: 30 miles
Distance to market: 90 miles

Not attractive at all :_____:_____:_____:_____ :_____:_____:_____ Very attractive
1 2 3 4 5 6 7

Section IV. General Information about YOUR MILL

1. What is the maximum distance (in miles) that you procure logs for this sawmill?

________ miles

2. How do you sell your finished lumber?

a. ☐ **FOB Mill**
b. ☐ **FOB Delivered**
c. ☐ **Both**

If you sell FOB delivered, what is the maximum distance (in miles) that you ship your finished lumber?

________ miles

3. Overall, do you believe that softwood sawmills tend to be located in geographical clusters or groups?

☐ YES ☐ NO ☐ NOT SURE

4. Based on your personal opinion do you think it is beneficial for softwood sawmills to be located close to each other in a cluster or group?

☐ YES ☐ NO ☐ NOT SURE

5. Do you expect to see an increase in softwood sawmill capacity in the U.S. South in the next 5 years?

☐ YES ☐ NO ☐ NOT SURE

6. Do you expect to see an increase in softwood sawmill capacity in the U.S. West in the next 5 years?

☐ YES ☐ NO ☐ NOT SURE

7. Do you expect to see an increase in softwood sawmill capacity in the Northeast region of the U.S. in the next 5 years?

☐ YES ☐ NO ☐ NOT SURE

8. Do you expect to see an increase in softwood sawmill capacity in the North Central region of the U.S. in the next 5 years?

☐ YES ☐ NO ☐ NOT SURE

Section V. Softwood Sawmill Clusters

Indicate your level of agreement with the following statements. (Circle only one for each)

1. By clustering together, softwood lumber mills have the following advantages over mills that are dispersed:

	Completely Disagree	Somewhat Disagree	Neither Disagree Nor Agree	Somewhat Agree	Strongly Agree
Better access to workers with managerial skills	1	2	3	4	5
Larger pool of skilled workers	1	2	3	4	5
Larger pool of unskilled labor	1	2	3	4	5
Better availability of raw materials	1	2	3	4	5
Better able to compete with other regions	1	2	3	4	5
Availability of more local suppliers	1	2	3	4	5
Potential collaboration among sawmills	1	2	3	4	5
Better access to information services	1	2	3	4	5
Greater opportunity to vertically integrate into manufacturing secondary products	1	2	3	4	5
Greater informal sharing of information between plants	1	2	3	4	5
Easier access to investment capital	1	2	3	4	5
Improved innovation through increased competition	1	2	3	4	5
A better organized industry	1	2	3	4	5

2. By clustering together, softwood lumber mills have the following disadvantages over mills that are dispersed:

	Completely Disagree	Somewhat Disagree	Neither Disagree Nor Agree	Somewhat Agree	Strongly Agree
Increased energy costs	1	2	3	4	5
Increased log prices	1	2	3	4	5
Increased labor costs	1	2	3	4	5
More congestion on local roads	1	2	3	4	5
Increased competition	1	2	3	4	5

Thank you for your cooperation and time in completing this survey!
Please return this survey by placing it in the postage paid envelope and dropping it in the nearest mailbox.

VITA

Francisco Xavier Aguilar Cabezas was born in Guayaquil, Ecuador, on August 17, 1977. After completing his high school degree at Javier High School he was awarded scholarships by the Vilaseca and W.K. Kellogg Foundations to pursue an engineering degree in agronomic sciences at Universidad Escuela de Agricultura de la Region Tropical Humeda (E.A.R.T.H.) in Costa Rica. He graduated from E.A.R.T.H. with honors and was named outstanding graduate receiving the University Medal in 1998. Following graduation Francisco volunteered for several months with a non-profit NGO supporting low-income communities in his native Guayaquil. In 1999 he was invited to participate in a multidisciplinary group to study the interaction of sustainable agriculture, community development and spirituality in communities in India and Sri Lanka. In Sri Lanka he promoted the use of biogas and organic farming. Francisco's work was recognized by the World Resources and the Earth Island Institutes. In October 2000 he began a master's degree in sustainable agricultural systems at the Royal Agricultural College (Cirencester, United Kingdom) as a scholar of the British Council. Upon graduation, in 2002 he joined the staff of the Office of the First Lady in the Ecuadorian Government. Francisco was hired as a specialist in sustainable farming systems to work for the Government-sponsored program on integrated community farming in provinces along the border between Ecuador and Peru. In 2003 he enrolled at Louisiana State University (LSU) to pursue a doctorate in forest products marketing and economics under Professor Richard P. Vlosky. In the spring of 2006 he was awarded a master's degree in agricultural economics with a concentration in natural resource and environmental policy. During his doctoral program at LSU he worked as an intern for Resources for the Future (RFF), a prominent non-profit and non-partisan think-tank that conducts independent research on environmental, energy, and natural resource economics issues.

Francisco has been the chief representative of the international student community at LSU during 2003-2004 and 2004-2005 as President of the International Student Association and President of the Board of the International Cultural Center. In 2006 Francisco was selected as one of three individuals nationwide to receive the 2006-2007 Joseph L. Fisher Dissertation Award to complete his doctoral research. This prestigious award, sponsored by RFF, is in recognition of Aguilar's research on determinants influencing the location of forest products manufacturers in Louisiana and the U.S. Francisco’s research efforts were recognized by the School of Natural Resources by making him a recipient of the 2007 Ben Stanley Award. In 2007 Francisco joined the Board of Directors of Envest, Inc., a non-profit organization based in Madison, Wisconsin, whose goal is to link microfinance and alternative credit to the formal financial sector and the international credit market in developing countries. The degree of Doctor of Philosophy will be awarded at the May 2007 Commencement.