ASSESSING REGIONAL VOLATILITY AND ESTIMATING REGIONAL COTTON ACRES IN THE UNITED STATES

by

BETH HOLMES

B.S., Arkansas State University, 1986

A THESIS

Submitted in partial fulfillment of the requirements

for the degree

MASTER OF AGRIBUSINESS

Department of Agricultural Economics

College of Agriculture

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2013

Approved by:

Major Professor
Vincent Amanor-Boadu

ABSTRACT

The objective of the research is to understand the volatility of cotton acres and estimate planted acres based on the factors that drive volatility in the United States at a regional level. Estimating cotton acres is important so that demand for cotton seed and technology can be anticipated and the appropriate investments in cotton seed production can be made.

Post Multi-Fiber Arrangement, the US cotton economy has entered a state of imperfect completion which makes cotton price, ending stocks and the relationship of cotton to other crops important in understanding volatility in cotton acres. Linear Regression, Random Forest and Partial Least Squares Neural Networks (PLS NN) were used to estimate cotton acres at a US and Regional Level. The modeling approaches used to estimate change in acres yielded similar performance for U.S. total, Southwest, and West. The PLS NN was slightly better for the Delta and Southeast, where more crop alternatives exist. Random Forest offered a different perspective on variable importance in all regions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The author wishes to acknowledge the Monsanto for supporting my pursuit of this degree. I also wish to express gratitude to my work colleagues, Huaiye Zhang, Saravanakumar Saminathan and Michael Lucido for their assistance and tolerance through this process. Appreciation also goes to Nanci Daesch and Tim Todd for their help and moral support through two and one half years of deadlines. Special gratitude goes to my son, Cy Stoltz, for his adaptability and acceptance of compromise in our day to day lives.

CHAPTER I: INTRODUCTION

The objective of the research is to understand the volatility of cotton acres in the United States so that demand for cotton seed and technology can be anticipated and the appropriate investments in production for cotton seed can be made. Demand impacts the volatility of planted cotton acres.

Monsanto's Delta and Pine Land Brand is a major supplier of cotton seed for production in the United States. Significant investments are made in the development of new technology, improved genetics, and seed production. As such, it's important to understand the reasons for change in cotton acres across years.

The history of cotton has been influenced by the location of the fiber consumption, policies and trade agreements for raw materials and textiles, government subsidies, demand and cost associated with synthetic fiber and the viability of alternative crops in the regions of production in the United States (U.S.). Utilization of government subsidized crop insurance, in lieu of government subsidized set aside, may have had an impact. Trade policies have changed over time, and a discussion of relevant policy issues will focus on the time period starting with the end of the multi-fiber arrangement (1994) to the present. These changes have had a direct impact of the existing market structure. The availability of adapted alternative crops has changed. Demand for those alternative crops is dynamic, as they also are commodity crops.

Cotton production is geographically spread across the U.S., and the factors influencing planted acres may not be affecting all regions equally. This study will examine potential drivers of cotton planted acres and propose statistical model to estimate acres by region and for the U.S. as a whole.

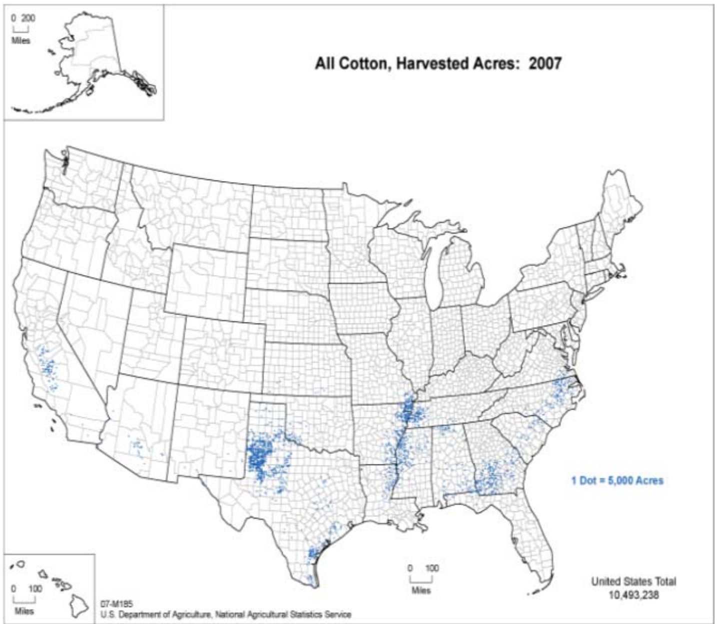
1.1 U.S. Cotton Acre Volatility

Cotton is grown as an annual crop in the United States. There are two principal types grown in the U.S.: Upland Cotton and Pima Cotton. Upland cotton production is concentrated in the southern states, with the largest cotton producing states being Texas and Georgia. Production in Texas is concentrated in the high plains. Texas alone accounted for more than 50% of total cotton acreage in the U.S. in 2010. The remainder is distributed across the southern states, with the majority of acreage in the Delta (along the Mississippi River.) These areas produce upland cotton. Pima cotton is grown in California, Arizona, New Mexico, and Texas. The following table shows the state breakout of acreage for cotton in 2012. The USDA notes cotton acreage has ranged between 9-15 million acres since the 1960's (USDA-NASS 2010).

Figure 1.1 shows the concentration of acres in 2007 as an example. Total cotton acres in this year were 12.83 million acre. The summary table shows the acres planted in 2012 from the largest to the smallest state.

Total cotton acres in the US have been fluctuating. Figure 1.2 shows total US acres for upland cotton planted acres. The minimum, maximum and standard deviation are included for the last thirty years (USDA-NASS 2010).

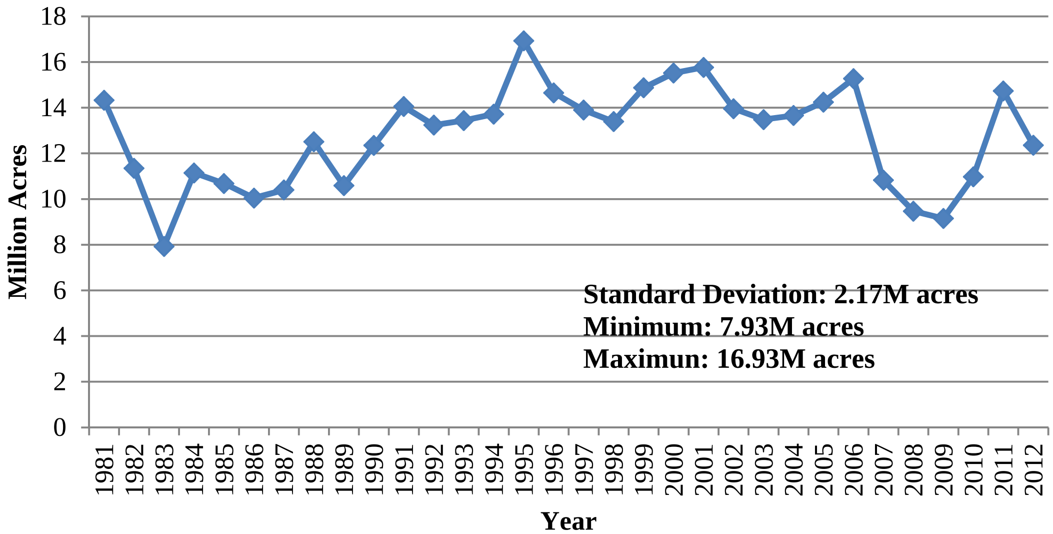
Figure 1.1: Example Annual Distribution of Cotton Planted Acres in the United States and Summary of 2012 Cotton Planted Acres by State



2012 Acres Planted(Acres)	Cotton planted acreage
TEXAS	6,608,000
GEORGIA	1,290,000
ARKANSAS	590,000
NORTH CAROLINA	585,000
MISSISSIPPI	470,000
TENNESSEE	380,000
ALABAMA	380,000
CALIFORNIA	367,000
MISSOURI	350,000
OKLAHOMA	305,000
SOUTH CAROLINA	298,000
LOUISIANA	230,000
ARIZONA	203,000
FLORIDA	108,000
VIRGINIA	86,000
KANSAS	57,000
NEW MEXICO	53,000
U.S. TOTAL ACRES	12,360,000

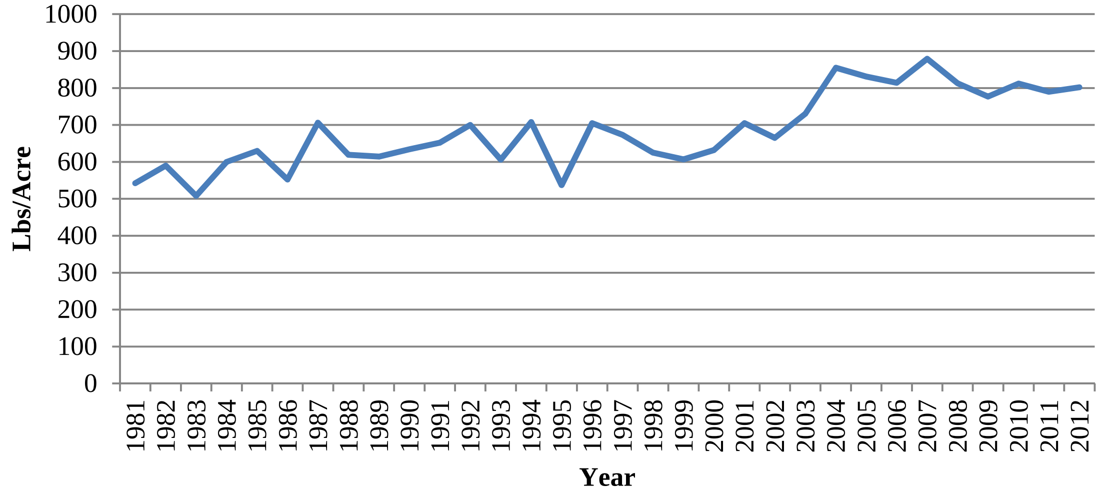
Source: (USDA-NASS 2010) (USDA- NASS 2012)

Figure 1.2: United States Cotton Acres in Millions of Acres from 1981 to 2012



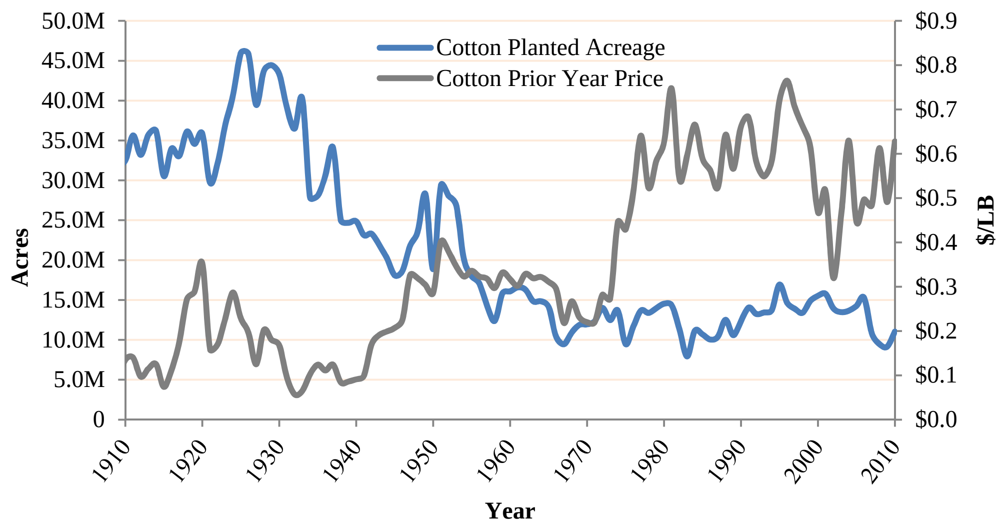
Yield has been increasing over time. Genetic improvements, biotechnology advancements and improved production management have contributed to yield. Relative to the early part of the 20th century acres have decreased and yet total production on the 8-17 million acres is on par with that achieved when the U.S. had acres greater than 30 million. (Figure 1.3) (USDA-NASS, 2010)

Figure 1.3: United States Cotton Acres in Millions of Acres from 1981 to 2012



Acreage has some impact on the cotton price as well. The graph below (Figure 1.4) shows US cotton acreage plotted relative to US farm gate price (USDA- NASS 2012). There is a positive correlation of 33% with cotton acreage and change in price from December to December. This correlation is lower relative to the same information for other US crops as seen in following table (Table 1.1).

Figure 1.4: Cotton Planted Acres in the United States versus Previous Year Cotton Farm Gate Price Over 100 Years



Source : (USDA- NASS 2012)

Table 1.1: Correlation of Acreage to Previous Year Crop Price in the United States for Corn, Soybean, Cotton and Wheat

Correlation			
Crop	***Period***	***Acreage vs. Prior Dec. Price***	***Acreage Change vs. Dec. Over Dec. Price Change Percentage***
Corn	***1970 - 2010***	***65%***	***57%***
Soy	***1924 - 2010***	***88%***	***34%***
Wheat	***1909 - 2010***	***-67%***	***39%***
Cotton	***1920 - 2010***	***22%***	***33%***

Source: (USDA- NASS 2012)

1.2 Monsanto's SEC Financial Summary (Cotton)

The financial summary below (Table 1.2) is an aggregate taken from Monsanto's SEC filings covering a time frame from the first cotton seed company acquisition to 2012. (Monsanto's fiscal year ends in August.) Monsanto's cotton business has generally shown increasing return over the years since 2005 with the exception of 2007 and 2012 where Gross Profit actually decreased in relationship to the previous year. Though not conclusive the decline in Gross Profit in these years could be attributed to decline in acres. The total cotton number of acres in 2007 was 10.83 million acres after a high of 15.27 million acres in 2006. Acres in 2012 were 12.36 million acres following 14.74 in 2011. This may suggest an issue when cotton acreage drops dramatically.

Table 1.2: Summary of Monsanto's SEC Filings Regarding Cotton from 2005-2012, With Calculated Growth Rate.

Dollars in Millions	2012	2011	2010	2009	2008	2007	2006	2005
Net Sales								
Cotton seed and traits	779	847	611	466	450	319	376	335
% Change from Previous Year	-8%	39%	31%	4%	41%	-15%	12%	
Gross Profit								
Cotton seed and traits	585	642	454	344	313	267	305	281
% Change from Previous Year	-9%	41%	32%	10%	17%	-12%	9%	
Growth Rate of Gross Profit		13.52%						
CAGR (0,3)		19.15%						
CAGR (0,7)		11.15%						

Source : (Monsanto 2012)

1.3 Monsanto's History in Cotton

The table (Table 1.3) below summarizes the technological achievements made from Monsanto's investment in cotton biotechnology and major milestones.

Table 1.3: Monsanto Milestones Associated with Cotton.

Year	Monsanto Cotton	Comments
1996	Bollgard® insect-protected cotton commercialized providing lepidopteron insect resistance	This is a Bt trait providing resistance to the major cotton pests of cotton bollworm, tobacco budworm and pink bollworm. This greatly reduced (or eliminated) the number of times a typical cotton farmer need to spray insecticide on his crop to prevent yield damage from insects. As in nearly all host:pathogen relationships, careful management is required to maintain durability of resistance.
1997	First Biotechnology Stacked Trait product released. Bollgard® and Roundup Ready® herbicide tolerance.	This allowed cotton farmers to use a cost effective way to manage weeds and enabled minimum tillage methods to be employed since tillage was no longer needed for weed control.
2002	Bollgard® cotton became the first biotech crop approved for India.	Adopted by 6 million farmers
2003	Introduced a second-generation trait product, Bollgard® II	This both expanded the range of insects controlled and increased the durability of resistance to targeted pests.
2005	Acquisition of the Stoneville cotton business, including its NexGen brand.	
2006	Release of second generation stack, Bollgard® II with Roundup Ready Flex®	Same insect protection benefits as Bollgard II. The improvement was in the herbicide tolerance by providing a longer, more flexible window for the use of Roundup Flex Technology.
	Bollgard® II released in India	
2007	Acquisition of Delta and Pine Land, divesture of Stoneville and NexGen	Corn hybrids adapted to the Delta cotton region become commercially available.
2010	Launch of a new herbicide for cotton and soybean acres called Warrant™ Herbicide	WARRANT™ herbicide is designed to be applied post-emergence of soybeans and cotton and pre-emergence to weeds. This herbicide provides 30 to 40 days of residual control of grasses and small seeded broadleaf weeds including the tough to control weeds, tall waterhemp and palmer amaranth.

(Monsanto 2012)

These achievements are a result of significant investments in research and product development as well as regulatory approvals. They include both technology milestones and acquisitions. There are also operational investments that are made on an annual basis for the development of new technology, genetic/variety development, and seed manufacturing costs. Ensuring that these investments are appropriately timed and sized will have a positive impact on the companies overall sustainability for positive growth.

CHAPTER II: LITERATURE REVIEW

2.1 Multi-Fiber Arrangement and Change in Market Dynamics

Market dynamic for cotton began to change when the Multi-Fiber Arrangement came to an end. This change impacted import export relationships as well as the location of textile manufacturing.

The Multi-Fiber Agreement (MFA) was an agreement that governed trade in textiles from 1974 through 2004. The purpose for this agreement was to allow a time period of adjustment for developed countries while developing nations began increasing production of wool, man-made fibers and cotton. The MFA was essentially a system of multilateral agreements establishing quotas that started with Japan and expanded over time to 40 countries. The intent was to protect textile workers in the developed nations The implementation saw the reassignment of many jobs to countries such as Bangladesh, Cambodia and El Salvador and more than 40 other countries, all with guaranteed market access through trade quotas with the US and EU. The boon to third world economies wasn't the original intent behind the MFA; it was following the direction of earlier quota policies initially set forward by the Kennedy administration to protect cotton production (Yearman 2005). While intended to protect producers and exporters, the protectionist MFA policy had a negative impact on the US consumer through increased clothing prices and consequently lower consumption.

The MFA did limit market access of the textile manufacturers from China and India during its duration. After the removal of MFA, cotton trade policy was governed by first the General Agreement on Trade Tariffs (GATT) and then by the World Trade Organization (WTO). The MFA expiration and the China's admission to the World Trade

Organization have changed China's position in the textile industry. (Wikipedia 2011) (GATT 1973).

The MFA was replaced by the Agreement on Textiles and Clothing (ATC) in 1995. This policy established a schedule for eliminating the MFA quotas and increasing growth rates in the import quantities still under quota. It lowered textile and soft goods (garment) tariffs and served as the impetus for bringing this trade sector under World Trade Organization (WTO) rules (MacDonald and Vollrath 2005).

The expiration of the MFA still has not resulted in free trade in cotton. The US has been in a dispute with Brazil over adherence to the WTO Doha Round of negotiations which would call for the US to amend its marketing loan program and counter cyclical crop payment programs that continue to subsidize US cotton production since 2003. Brazil won this ruling in 2010 and the US has since provided recompense to Brazil, both monetarily, with payments of $147 million, and by allowing some import of Brazilian beef into the US. Models from the International Centre for Trade and Sustainable Development indicate that if the US abided completely by the Doha WTO round, world market price would increase as much as 2% and the US could see a reduction of cotton acres of 6-10%. While adherence would definitely impact world price it's difficult to know if the long term impact would mean higher prices and more stability over the long term. The US did pass a new farm bill in 2008 but it did little to address reforms for the cotton subsidies that keep price artificially low (ICTSD 2010).

The chart (Figure 2.1) below supports the misalignment of the US to the other major cotton producing countries (Brazil, China, India and Pakistan) through the expression of ending stocks expressed as percent change annually. A look at the volume in ending stocks clearly demonstrates a picture of changing influence in ending stocks concentration from China (Figure 2.2).

Figure 2.1: Percent Change in U.S. Cotton Ending Stocks by Year Compared to the Percent Change of the Sum in Cotton Ending Stocks for Brazil, China, India and Pakistan from 1970-2012.

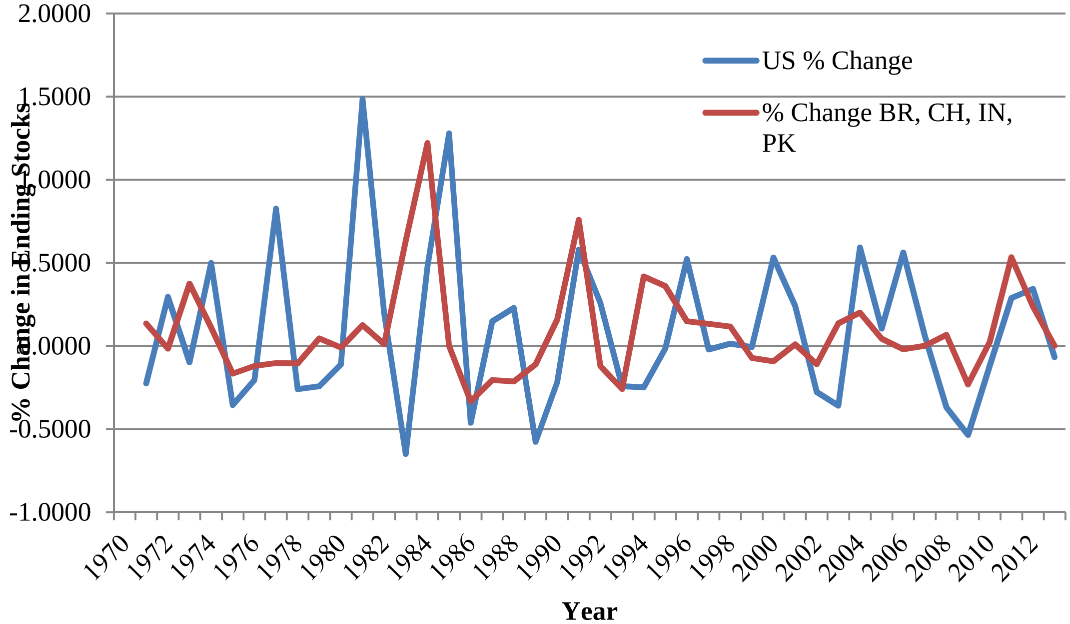
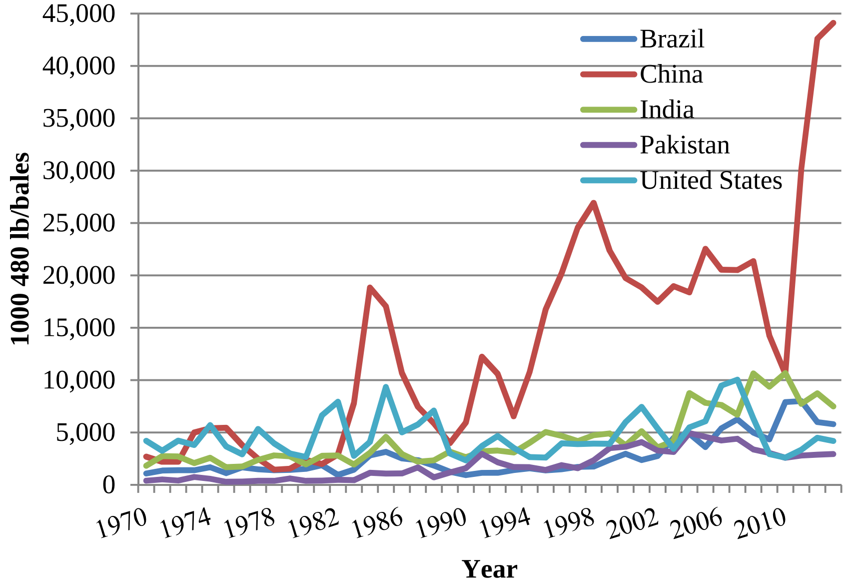


Figure 2.2: Total Cotton Ending Stocks for Brazil, China, India, Pakistan and the United States from 1970-2012.



A relationship between fiber consumption and Gross Domestic Product (GDP) was proposed by Baffes and Gohu (2005) based on data from many countries with various stages of economic development. Their study showed a reasonably linear relationship between GDP and fiber consumption in the year 2000. Fiber consumption was highest among developed nations. Disruptions to long term trend GDP could therefore have a negative impact on consumption and consequently demand.

The United States is the third largest cotton producer following China and India. With the removal of trade restrictions in the multi-fiber agreement, textiles from China have become more available. U.S. cotton exports have nearly doubled from 1997-2007 with 13.6 million 480-pound bales exported in 2007 while domestic mill use declined by more than 50% from 11.3 bales to 4.6 bales in the same time frame. This links the U.S. cotton economy closely to foreign demand, specifically import demand from China. Thus, China has a high concentration of buying power with its dominance of the textile industry (Pan, Hudson and Ethridge 2010) . Further, this supports the proposition that the international cotton market is one of imperfect competition. While there is not consensus on the effect of developing countries, it is generally thought this market structure will have more impact on global cotton prices than U.S. crop subsidies programs alone. Hudson et al. (2011) have projected this relationship of the U.S. as the largest cotton exporter and China as the largest importer to persist for the next decade and base this on a more favorable exchange rate between the Chinese Yuan and the U.S. Dollar. (Hudson, et al. 2011).

According to a study by Isengildina-Massa and MacDonald (2009), structural changes in the market suggest that change in U.S. cotton price will be most influenced by U.S. cotton supply, U.S. stocks-to-use ratio, China's net imports as a share of world consumption, the proportion of U.S. cotton in the loan program, and the world supply of cotton. These variables explain 68% of the cotton price variation from 1974-2007. There is nearly a 1:1 relationship where a 1 percent increase in US supply in the previous year causes a 0.9 percent drop in U.S. cotton price. Changes in foreign supply affect U.S. prices on a nearly one-to-one basis as well, with prices falling as foreign supply rises.

2.2 Risk Mitigation Approaches

The cotton trade relationship with China highlights the importance of risk mitigation strategies for cotton producers (Foreman 2012). Most cotton producers can produce other commodities as a risk mitigation strategy. Foreman's research separated cotton growers by low cost, mid-cost, and high cost in 2007. Foreman noted that the number of commodities produced averaged between 3.1-4.2 depending on the location and cost structure of those producers. The southwest had fewer alternatives than the Delta and Southeast. Corn, soybean and wheat are all alternatives in Delta and Southeast, whereas wheat is the most prevalent alternative in the Southwest and West. Corn is grown in these regions but not on an equivalent number of acres. Where alternatives are possible, the net return of each commodity is considered before a planting decision is made, according to Foreman's study.

Crop insurance is used as another risk mitigation strategy. Ninety-five percent of cotton acres in 2007 were covered by crop insurance. As part of the same study, Foreman found that there was a difference in the type of insurance sought by growers. Southwest growers are more likely to pursue Federal revenue insurance. These farms are more likely to be 1500 acres or more. The very large producers of the Delta with gross sales totaling a $1M or more were more likely to mitigate risk with buy up insurance that provides coverage for low yield but not price as revenue insurance does. The Southwest has fewer commodity diversification options than the Delta and the limited options are likely the reason of differences in approach to crop insurance.

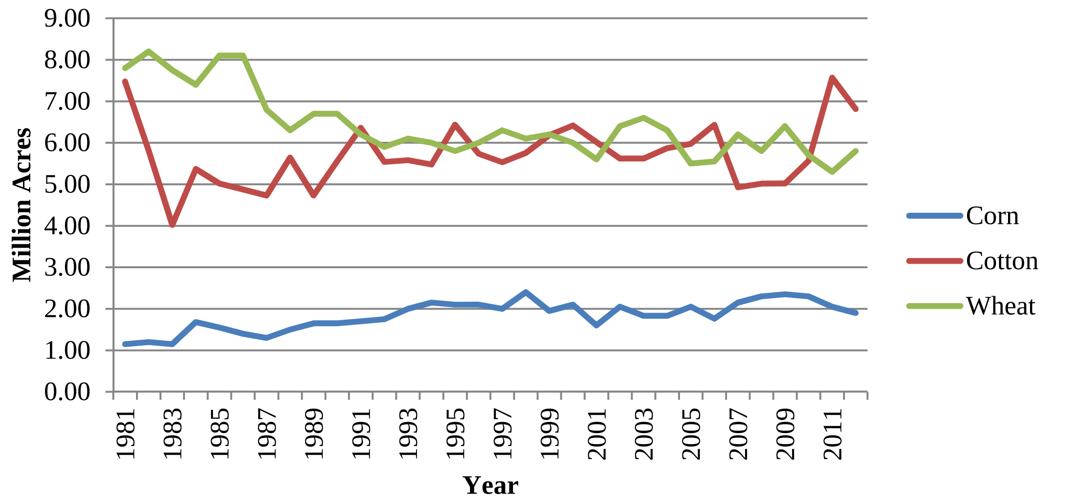
Lack of continuity in government programs has been another source of variability affecting cotton planted area in the U.S. Cotton acres in the southeast US have shown responsiveness to the cotton loan rate, deficiency payments, and payment in kind programs. Recently, Foreman (2012) noted that the southwest region is more likely to engage in government supported revenue insurance programs.

Alternative crops and improved management practices may have been elevated in importance with the enactment of the Federal Agricultural Improvement and Reform (FAIR) Act in 1996. Under FAIR, producers must be more responsive to available market and production information, because many government price supports have been eliminated. (Hudson, et al. 2011)

2.2.1 The Relationship of Alternative Crops to Cotton Acres

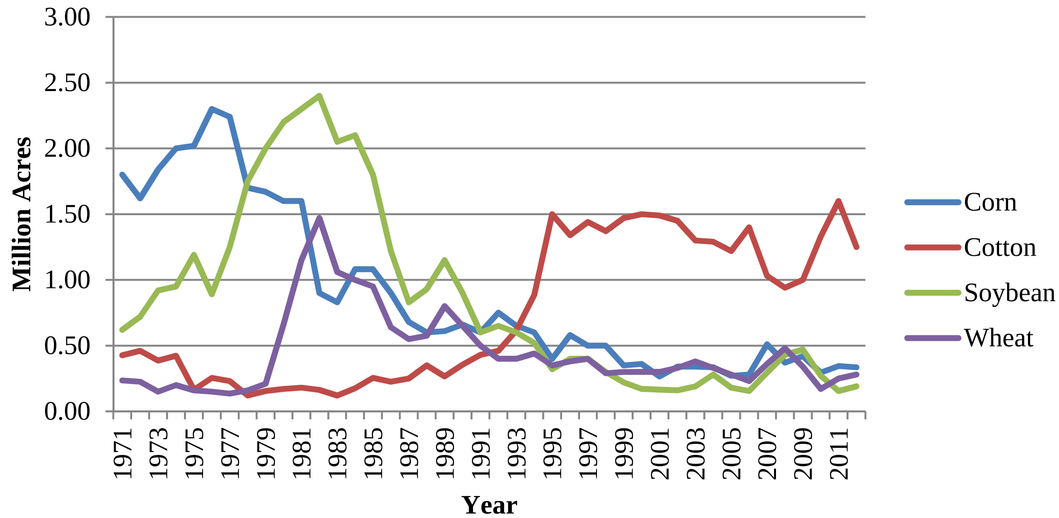
The ability to grow alternative crops may also influence the number of acres planted to cotton in any given year. This will be dependent on the crops that can be grown in that state based on adaptability and infrastructure. Several key relationships are highlighted below. Texas is the largest state in the Southwest region (Texas, Oklahoma and Kansas) and Georgia is the largest cotton producing state in the Southeast (Georgia, Alabama, North Carolina, and South Carolina.) Combined these two states account for 70% of the cotton acres in the U.S. The charts below (Figure 2.3 and Figure 2.4) show select and interesting relationships to other crops in Texas and Georgia. Texas shows a possible negative relationship between cotton and wheat as well as increasing corn acres over time. Georgia shows a much different relationship. A look back at the Monsanto timeline and the introduction of varieties and hybrids with biotechnology would suggest that this influenced the growth of cotton at the expense of other crops.

Figure 2.3: Texas Planted Acres for Cotton, Winter Wheat and Corn, 1981-2012.



Source: (USDA- NASS 2012)

Figure 2.4: Georgia Planted Acres for Cotton, Winter Wheat , Soybean and Corn, 1971-2012.



Source: (USDA- NASS 2012)

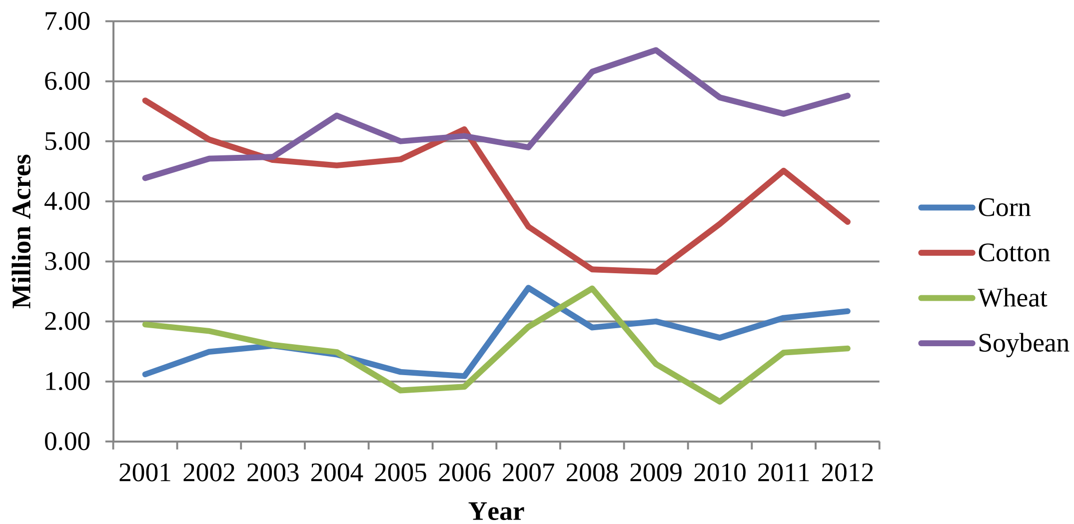
Changes in production efficiency have caused a shift in crop acres in the southeast. The Boll Weevil Eradication (BWE) plan in Georgia reduced the prevalence and the need for pesticides for control of this difficult pest, making cotton a much more profitable crop to grow. Tribble et al. (1999) credit the BWE program with a magnitude increase of cotton acres from the early 1980’s at a historic low of 120,000 acres to 1.5 million acres by 1996.

The introduction of biotechnology created production efficiency changes across crops. This was true for cotton, corn and soybean alike. Liang, et al note that adoption of genetically engineered crops accounted for 86% of all upland cotton, 80% of corn and 92% of all soybean in the U.S. by 2007 according to data collected by USDA-ERS. While seed cost have increased, production efficiency is gained through the reduction in other input costs, higher yields and less variability. (Liang, et al. 2011) The impact of adapted corn hybrids with biotechnology is also visible when looking at the increase in corn acreage can be seen in the Delta region (Arkansas, Missouri, Louisiana, Mississippi, and Tennessee).

The chart below illustrates this effect over time for Arkansas and Mississippi (Figure 2.5). Cotton acres are declining and corn acres are increasing.

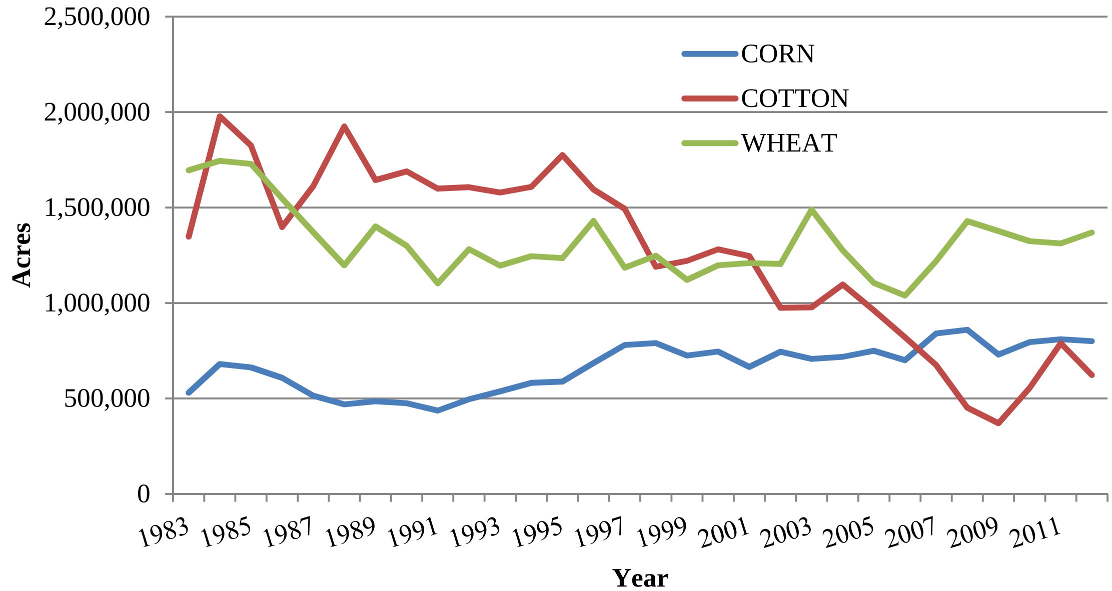
California and the other western states show a downward cotton acres trend but these states produce mostly Pima cotton. These states are also showing a decline in cotton acres relative to a slight increase in corn (Figure 2.7). Production of this type of cotton may be shifting to one of the other cotton producing countries.

Figure 2.5: Arkansas and Mississippi Planted Acres for Cotton, Winter Wheat, Soybean and Corn, 2001-2012.



(USDA- NASS 2012)

Figure 2.6: California, New Mexico and Arizona Planted Acres for Cotton, Winter Wheat and Corn, 1971-2012.



(USDA- NASS 2012)

In 2011, Liang, et al estimated supply elasticity values for corn, cotton, and soybeans of approximately 0.670, 0.506, and 0.195, respectively.

Cotton is more price elastic than the potential substitute of soybean. Corn and cotton respond more to fluctuations in price and may thus be competitive for acres based on the relative relationship of price.

Crop insurance has been shown to influence planted acres but has less impact than returns. Premium changes since 1994 have made crop insurance even more affordable. Tronstad and Bool (2010) estimated that a 1% increase in expected net returns would increase cotton acreage by 0.035%, while a 1% increase in expected net market returns would increase cotton acreage by 0.222%. The portion of total subsidy support going to cotton has increased since 1995, averaging 11.5% for the time period, 1995-2005. While the expected market returns may have more influence, crop insurance is still an important part of the equation. While the expected market returns may have more influence, crop insurance is still an important part of the equation (Foreman 2012). In earlier work, Parrott and McIntosh observed that cash price was more influential than government programs.

Cotton acreage is influenced by net expected returns. Farmers take cost of production and expected market price into account when deciding on what to plant. As alternative crops become more profitable, acres in even traditional cotton growing areas such as the Delta can shift. Mutuc and Hudson (2010) have noted a sustained decline in cotton acres since 2006. This coincides with increased production of bio-fuels and higher returns for corn. Growers have shown some reluctance to switch to from cotton to corn, even forgoing potentially higher returns, due in part to capital investments made in equipment for cotton production. Nonetheless, cotton acres declined in three consecutive years from 2006-2008. As corn prices started to rise and the availability of suitable hybrids increased, the likelihood that farmers would substitute corn for cotton when making planting decisions increased. This demand was influenced by the Renewable Fuels Standard which ensures that gasoline marketed in the U.S. contains a designated percentage of ethanol (Mutuc and Hudson 2010). Fannin and Paxton (2011) show that while cotton producers were willing to forego revenue for a time, even when the foregone return was as high as 35%, a tipping point exists. The 2007 price of corn, for example, helped to convert Delta cotton producers to corn at the expense of cotton. A similar spike for cotton with a drop in corn and soybean price would be required to create an upward progress for cotton acres.

In summary, many factors may exert influence on cotton acres and add to the volatility of this crop. A selected set of these variables will be examined to determine their impact on predicting cotton acreage, including previous year's cotton acres, change in cotton futures price, corn, the difference between cotton futures price and alternative crop futures price (corn and soybean), ending stocks (corn, soybean and wheat relative to cotton), previous years economic growth indicators for China, the largest importer (Gross Domestic Product and the US the largest finished goods consumer (Gross Domestic Income and Consumer Price Index) and finally price of ethanol, a major demand driver for corn, a key alternative crop to cotton.

CHAPTER III METHODS AND RESULTS

This chapter contains the theoretical model, detailed descriptions of the variables used to represent that model, an assessment of those variables for each region as well as the U.S. overall, model performance for each region and the U.S. and finally results and discussion by region and for the U.S.

3.1 Theoretical Model and Variable Definition

Analysis was performed at both the U.S. and Regional level. The theoretical model proposed is:

$$A_{it} = f\,(A_{it-1}, D_{jt}, S_{jt}, C_{t-1}, B_{jt-1})$$

Where A_{it} is the change in cotton acres in region *i* in period t where *i*=Southeast, West, Southwest, Delta and the U.S. *D* is the relative price between cotton and alternative crops, S is the relative supply between cotton and alternative crops, *C* is the growth rate in cotton consumption and *B* is the consumption growth in alternative crops. The subscript *j* represents the alternative crops (wheat, corn and soy). The relative price is defined as:

$$D_{ijt} = P_{fc} \,/\, P_{fj}$$

Where P_f is the futures price and *c* is cotton and *j* is the other crops of interest. This is represented by the futures price of cotton and the alternative futures price for corn, soybean and wheat, represented as ratios. Relative supply, S is defined as:

$$S_{ijt}=E_{fc}/E_{fj}$$

Where E_f is the ending stock and *c* is cotton and *j* is the other crops of interest. Relative Supply is represented by ending stocks for cotton and the alternative crops of corn, soybean and wheat, represented as ratios.

The growth rate for cotton consumption, C, is expected to be a leading indicator (t-1) and is represented by previous year U.S. Gross Domestic Income (GDI) as a measure of consumption potential and by previous year Consumer Price Index (CPI) as a measure of inflation. Growth in consumption is also represented by import demand from China but indirectly as previous year annual change in Gross Domestic Product (GDP) for China. The decision to use China GDP rather than China cotton imports was an attempt to identify factors which might serve as leading indicators in a twelve to twenty-four month time horizon. The concern was that estimates would not be available for China cotton imports. US GDP was also considered and discarded after data exploration for the U.S. level cotton acre estimations.

The consumption growth for alternative crops, B, is also considered as a leading indicator (t-1) where j is for the other crops of interest. The last variable considered was lag ethanol price as an indicator for alternative crop (corn) demand.

This set of analyses is considered foundational to future work; all variables were assessed for all regions. The acres themselves are regional; values for all other variables remain the same across regions. Table 3.1 outlines the variables and their sources.

Table 3.1: Summary of Variables Used for Estimating Cotton Acres at the U.S. and Regional Level and Associated Sources.

Variable	Description	Source
Change in Acres	Dependent variable. Percent change in cotton acres since previous year	Calculated
Change	Change in cotton acres at U.S. level or Regional level.	USDA
Lag_ Acres	Previous year cotton acres at U.S. level or Regional level.	USDA
Cotton_Change_FPDEC	Change in the December futures price of the next year contract from the previous year.	CBOT Calculated
Cotton_Corn_FPRDEC Cotton_Soy_FPRDEC	December price of the next year contract. Future Price Ratios. Ratio of future prices of cotton to corn and soybean. (December contract for corn and cotton and November contract for soy)	CBOT Calculated
Cotton_Corn_ESR Cotton_Wheat_ESR Cotton_Soy_ESR	Ending Stocks Ratios. Corn, Soybean and Wheat; USDA September Quarterly Ending Stock (bushels). Cotton; Annual Ending Stocks (1000 480lb bales)	USDA Calculated
Lag_U.S._GDI	Previous year U.S. Gross Domestic Income (a component of GDP)	U.S. Bureau of Economic Analysis

Lag_U.S._ CPI	Previous year U.S. Consumer Price Index	U.S. Bureau of Economic Analysis
Lag_ China_GDP	Previous year growth % in China Gross Domestic Product	Global Insights
Lag_Ethanol	Ethanol Price, FOB Omaha, cal. yr., United States Units of Measure: dollars per gallon Historic Data Source: RFA Data Edge: 2011 Forecast Date: November 2012	Global Insights
Other Descriptors:		
Delta Region	Missouri, Arkansas, Mississippi, Lousiana and Tennessee	
Southeast Region	Alabama, Georgia, North Carolina, South Carolina	
Southwest Region	Texas, Oklahoma, Kansas	
West Region	California, New Mexico, Arizona	
Years	1983-2012	

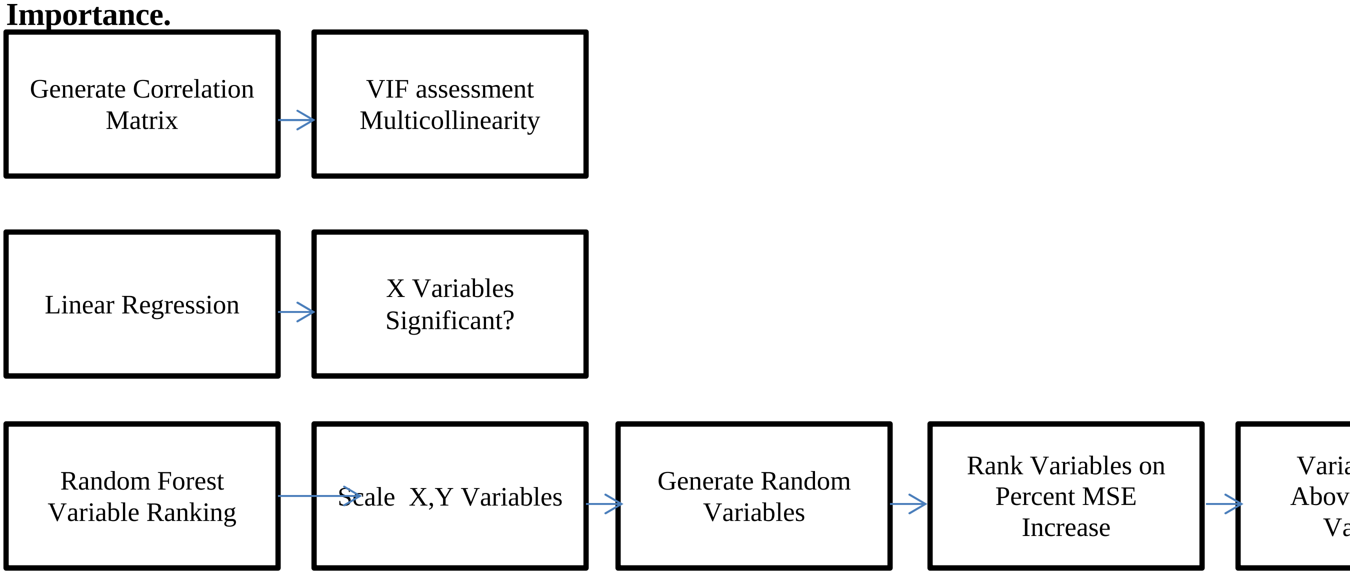
3.2 Variable Assessment

Variables were assessed for correlation, multicollinearity, significance through linear regression and rank using Random Forest for each region and for the U.S. A correlation matrix is generated and variance inflation analyses are run to determine if multicollinearity exists using a threshold of 5.00%. The presence of multicollinearity must be considered when assigning importance to the independent variables impacting variance. Linear regression is performed to estimate the coefficients associated with each variable.

Random Forest is performed to achieve a variable ranking. Random Forest analyses aggregate the response across many decision trees, where each tree is similar to an un-pruned Classification and Regression Tree (CART) analysis (Ho 1995). An advantage of this methodology is that it outputs variance importance measures. This analysis defines the tree number = 200. Mean Square Error is generated and the percent change associated with that variable is used to rank the variables. Further, the ranks are averaged over 500 runs. Variables were considered important if they occurred in rank above the first random variable. Potential pitfalls with this method were addressed by first scaling the variables prior to analysis (Strobl, et al. 2007).

The general process flow for variable assessment is outlined in Figure 3.1. Assessments by region will be discussed in the subsequent sub sections.

Figure 3.1: Analysis Flow Chart for Estimating Change in Cotton Acres and Variable Importance.



3.2.1 Variable Assessment for the U.S.

The correlation matrix indicates a strong negative (<-0.50) relationships between the dependent variable, change in U.S. acres and the cotton to corn ending stocks ratio and the cotton to wheat ending stocks ratio (Table 4.1).

There are also interesting relationships between independent variables, which indicated multicollinearity. Since the values for these variables don't change from region to region, the relationships persist across regions and will reviewed only once in this section.   The correlation matrix indicates strong positive relationships (>0.50) between; the cotton to corn ending stock ratio and the cotton to wheat ending stocks ratio, cotton to soybean ending stock ratio and cotton wheat ending stock ratio, the cotton to soybean future price ratio and the cotton to corn ending price ratio, previous year ethanol price and previous year U.S. GDI. Strong negative correlations exist between previous year US GDI and cotton to corn future price ratio, previous year ethanol price and cotton to corn future price ratio, previous year U.S. CPI and previous year U.S. GDI, cotton to soybean futures price and cotton to soybean ending stock ratio and cotton to soybean futures price and previous year U.S. GDI (Table 3.2).

Table 3.2: U.S. Level Correlation Matrix

	CottonCorn_ESR	Cotton_Wheat_ESR	Cotton_Change_FPDEC	Cotton_Soy_ESR	Lag_Cotton_Acres	Cotton_Corn_FPRDEC	Lag_US_GDI	Lag_US_CPI	Lag_Ethanol	Cotton_Soy_FPRDEC	Lag_China_GDP	y
Cotton_Corn_ESR	1.00											
Cotton_Wheat_ESR	0.59	1.00										
Cotton_Change_FPDEC	0.01	-0.16	1.00									
Cotton_Soy_ESR	0.49	0.60	-0.11	1.00								
Lag_Cotton_Acres	0.41	0.19	-0.21	-0.05	1.00							
Cotton_Corn_FPRDEC	-0.40	-0.46	0.15	-0.48	0.05	1.00						
Lag_US_GDI	0.35	0.34	0.19	0.36	0.23	-0.73	1.00					
Lag_US_CPI	-0.01	0.12	-0.40	0.00	-0.04	0.14	-0.50	1.00				
Lag_Ethanol	0.37	0.42	-0.07	0.32	-0.17	-0.79	0.63	0.07	1.00			
Cotton_Soy_FPRDEC	-0.25	-0.43	0.04	-0.50	0.28	0.88	-0.66	0.14	-0.76	1.00		
Lag_China_GDP	0.30	0.32	0.06	0.14	-0.05	-0.11	-0.03	-0.11	0.21	-0.17	1.00	
y	-0.55	-0.58	0.39	-0.42	-0.47	0.28	-0.09	-0.32	-0.25	0.13	-0.17	1.00

The variance inflation analysis assesses the impact of multicollinearity of theses variables on the change in U.S. cotton acres (Table 3.3). Previous year cotton acres, the cotton to corn futures price ratio, previous year U.S. GDI, previous year ethanol price, and the cotton to soybean futures price ratio all exceed the 5% threshold.

Table 3.3: U.S. Percent Change in Linear Regression Variance Inflation Factors

Variable	VIF
Cotton_Corn_ESR	3.54
Cotton_Wheat_ESR	2.35
Cotton_Change_FPDEC	2.61
Cotton_Soy_ESR	3.01
Lag_Cotton_Acres	6.57
Cotton_Corn_FPRDEC	8.09
Lag_US_GDI	16.73
Lag_US_CPI	4.17
Lag_Ethanol	9.14
Cotton_Soy_FPRDEC	7.78
Lag_China_GDP	1.94

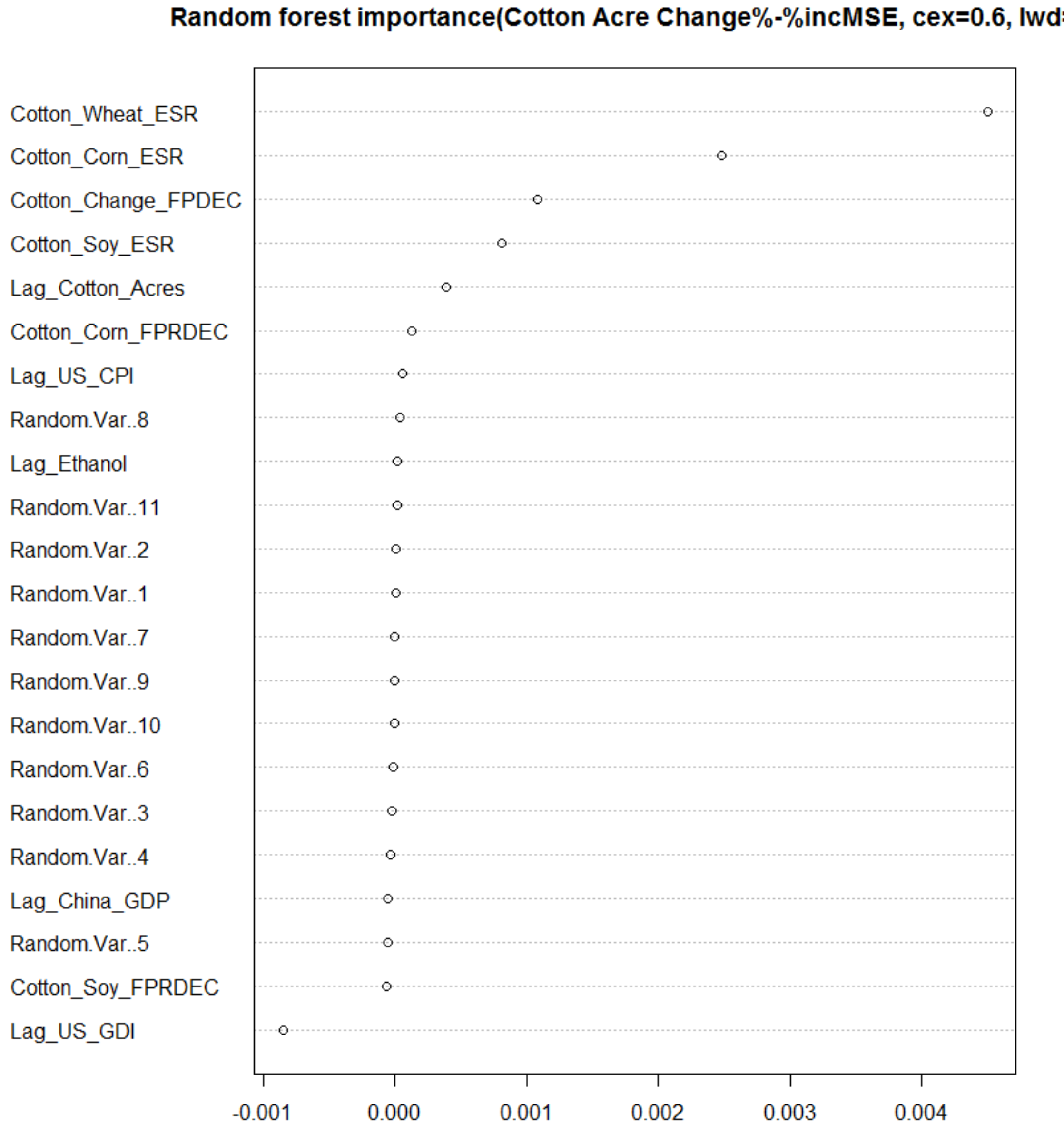
Linear regression was performed, the R square was 0.71 and the adjusted R square was 0.54, and the variable coefficients, standard errors, t statistic and P value are summarized below in Table 3.4. Previous year cotton acres are significant at the 0.01 level of significance and previous year U.S. GDI and the cotton to soybean ending stock ratio is significant at the 0.05 level.

Table 3.4: Summary of Linear Regression Independent Variables in Response to Percent Change in U.S. Cotton Acres

Variable	Coefficients	Standard Error	t Stat	P-value
Intercept	0.614686	0.379687	1.618928	0.122851
Cotton_Corn_ESR	0.018608	0.025772	0.722025	0.479555
Cotton_Wheat_ESR	-0.03066	0.02637	-1.16264	0.260157
Cotton_Change_FPDEC	-0.13473	0.176061	-0.76525	0.454039
Cotton_Soy_ESR	-0.00729	0.003316	-2.20003	0.041106
Lag_Cotton_Acres	-7.1E-08	2.36E-08	-3.02671	0.007251**
Cotton_Corn_FPRDEC	0.89489	1.199647	0.745961	0.465322
Lag_US_GDI	5.09E-05	2.2E-05	2.310612	0.032913*
Lag_US_CPI	0.02757	0.033162	0.831363	0.416661
Lag_Ethanol	-0.26951	0.132677	-2.03133	0.057255
Cotton_Soy_FPRDEC	0.648014	3.341684	0.193918	0.848411
Lag_China_GDP	0.008243	0.010612	0.776809	0.44736

The Random Forest variable ranking shows a different view of the variable contribution to the model (Figure 3.2). The variable rank of those s with a percent MSE greater than the first random variable is, in order; cotton to wheat ending stock ratio, cotton to corn ending stock ratio, change in cotton futures price, cotton to soybean ending stock ratio, previous year cotton acres, cotton to corn futures price ratio, previous year cotton acres and the previous year U.S. CPI.

Figure 3.2: U.S. Random Forest Percent Increase in Mean Square Error



3.2.2 Variable Assessment for the Delta Region

The correlation matrix indicates a strong negative (<-0.50) relationships between the dependent variable, change in Delta acres and the cotton to corn ending stocks ratio, the cotton to wheat ending stocks ratio, and the cotton to soybean ending stocks ratio (Table 3.5).

Table 3.5: Delta Region Correlation Matrix

	Lag_Delta_Acres	Cotton_Corn_ESR	Cotton_Wheat_ESR	Cotton_Change_FPDEC	Cotton_Soy_ESR	Cotton_Corn_FPRDEC	Lag_US_GDI	Lag_US_CPI	Lag_Ethanol	Cotton_Soy_FPRDEC	Lag_China_GDP	y
Lag_Delta_Acres	1.00											
Cotton_Corn_ESR	0.36	1.00										
Cotton_Wheat_ESR	0.22	0.59	1.00									
Cotton_Change_FPDEC	-0.17	0.01	-0.16	1.00								
Cotton_Soy_ESR	-0.10	0.49	0.60	-0.11	1.00							
Cotton_Corn_FPRDEC	0.28	-0.40	-0.46	0.15	-0.48	1.00						
Lag_US_GDI	-0.08	0.35	0.34	0.19	0.36	-0.73	1.00					
Lag_US_CPI	0.08	-0.01	0.12	-0.40	0.00	0.14	-0.50	1.00				
Lag_Ethanol	-0.41	0.37	0.42	-0.07	0.32	-0.79	0.63	0.07	1.00			
Cotton_Soy_FPRDEC	0.41	-0.25	-0.43	0.04	-0.50	0.88	-0.66	0.14	-0.76	1.00		
Lag_China_GDP	0.05	0.30	0.32	0.06	0.14	-0.11	-0.03	-0.11	0.21	-0.17	1.00	
y	-0.33	-0.62	-0.62	0.27	-0.52	0.38	-0.24	-0.21	-0.34	0.28	-0.23	1.00

The variance inflation analysis assesses the impact of multicollinearity of theses variables on the change in Delta cotton acres (Table 3.6). Previous year cotton acres, the cotton to corn futures price ratio, previous year U.S. GDI, previous year ethanol price, and the cotton to soybean futures price ratio all exceed the 5% threshold.

Table 3.6: Delta Percent Change Linear Regression VIF

X Variable	VIF
Lag_Delta_Acres	6.88
Cotton_Corn_ESR	4.03
Cotton_Wheat_ESR	2.76
Cotton_Change_FPDEC	2.21
Cotton_Soy_ESR	3.57
Cotton_Corn_FPRDEC	8.46
Lag_US_GDI	12.57
Lag_US_CPI	4.50
Lag_Ethanol	12.97
Cotton_Soy_FPRDEC	6.50
Lag_China_GDP	2.05

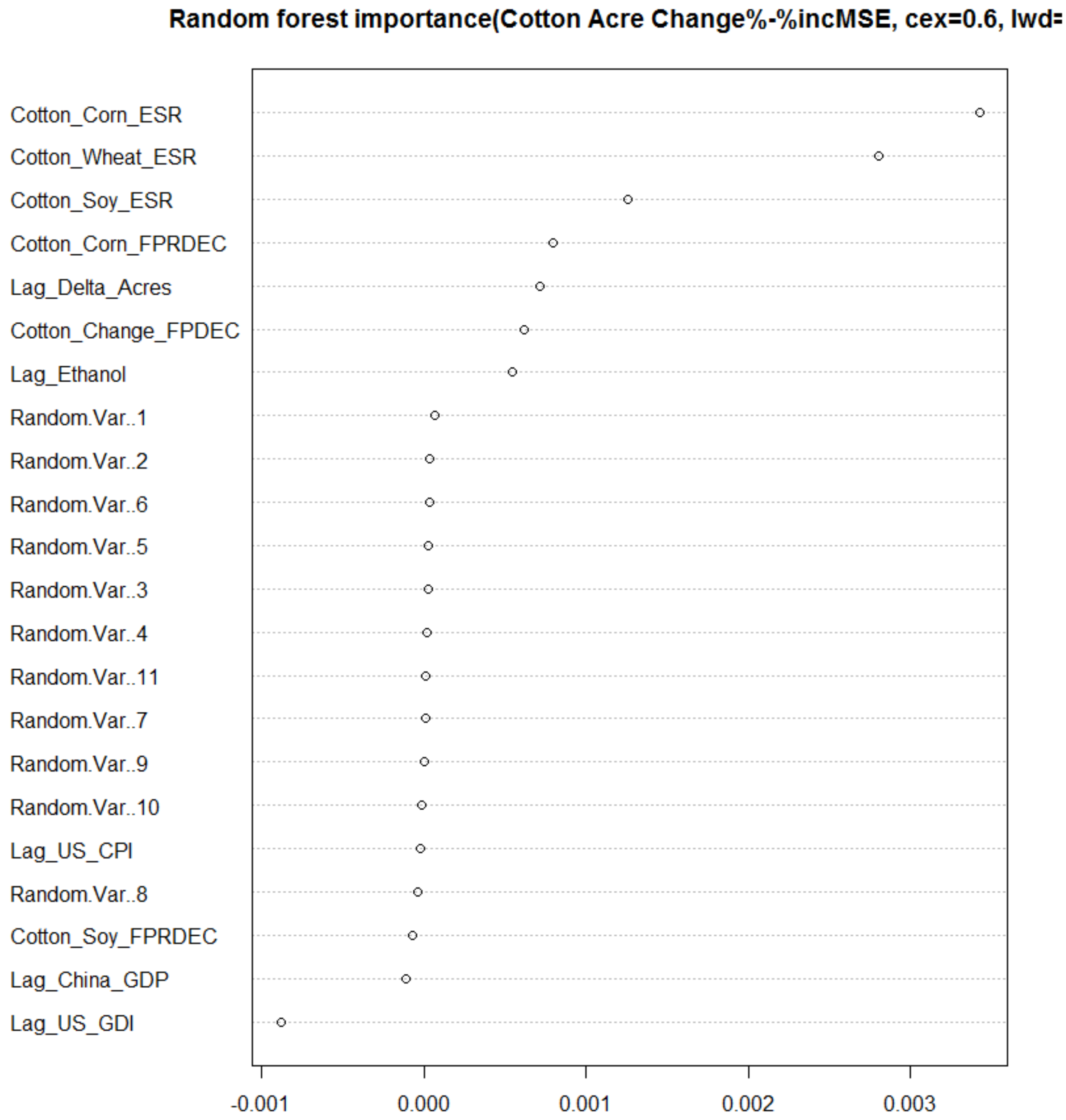
The Linear regression for the Delta had an R square was 0.60 and an adjusted R square was 0.36. The variable coefficients, standard errors, t statistic and P value are summarized below in Table 3.7. There are no variables that are significant.

Table 3.7: Summary of Linear Regression Independent Variables in Response to Percent Change in Delta Cotton Acres

Variable	Coefficients	Standard Error	t Stat	P-value
Intercept	0.6701	0.5375115	1.2467	0.2285
Cotton_Soy_ESR	-0.0074	0.0050887	-1.4448	0.1657
Lag_Delta_Acres	0.0000	0.0000001	-1.3208	0.2031
Lag_Ethanol	-0.2030	0.2224796	-0.9126	0.3735
Lag_US_GDI	0.0000	0.0000269	0.6039	0.5534
Cotton_Corn_FPRDEC	0.7612	1.7276245	0.4406	0.6647
Cotton_Wheat_ESR	-0.0119	0.0402288	-0.2963	0.7704
Cotton_Corn_ESR	-0.0100	0.0386930	-0.2588	0.7987
Cotton_Soy_FPRDEC	-1.0376	4.3026564	-0.2412	0.8122
Lag_China_GDP	0.0012	0.0153805	0.0777	0.9389
Lag_US_CPI	0.0035	0.0485021	0.0712	0.9440
Cotton_Change_FPDEC	0.0134	0.2280170	0.0588	0.9537

The Random Forest variable ranking shows several variables appearing before the first random variable (Figure 3.3). The variables in rank order are: cotton to corn ending stock ratio, cotton to wheat ending stock ratio, cotton to soybean ending stock ratio, the cotton to corn futures price ratio, previous year Delta, change in cotton futures price and previous year ethanol price.

Figure 3.3: Delta Region Random Forest Variable Ranking



3.2.3 Variable Assessment for the Southeast Region

The correlation matrix indicates a strong negative (<-0.50) relationships between the dependent variable, change in Southeast cotton acres and the cotton to corn ending stocks ratio, the cotton to wheat ending stocks ratio, and the cotton to soybean ending stocks ratio (Table 3.8).

Table 3.8: Southeast Region Correlation Matrix.

	Lag_Southeast_Acres	Cotton_Corn_ESR	Cotton_Wheat_ESR	Cotton_Change_FPDEC	Cotton_Soy_ESR	Cotton_Corn_FPRDEC	Lag_US_GDI	Lag_US_CPI	Lag_Ethanol	Cotton_Soy_FPRDEC	Lag_China_GDP	y
Lag_Southeast_Acres	1.00											
Cotton_Corn_ESR	0.45	1.00										
Cotton_Wheat_ESR	0.23	0.59	1.00									
Cotton_Change_FPDEC	0.00	0.01	-0.16	1.00								
Cotton_Soy_ESR	0.17	0.49	0.60	-0.11	1.00							
Cotton_Corn_FPRDEC	-0.27	-0.40	-0.46	0.15	-0.48	1.00						
Lag_US_GDI	0.69	0.35	0.34	0.19	0.36	-0.73	1.00					
Lag_US_CPI	-0.46	-0.01	0.12	-0.40	0.00	0.14	-0.50	1.00				
Lag_Ethanol	0.11	0.37	0.42	-0.07	0.32	-0.79	0.63	0.07	1.00			
Cotton_Soy_FPRDEC	-0.04	-0.25	-0.43	0.04	-0.50	0.88	-0.66	0.14	-0.76	1.00		
Lag_China_GDP	-0.12	0.30	0.32	0.06	0.14	-0.11	-0.03	-0.11	0.21	-0.17	1.00	
y	-0.37	-0.58	-0.62	0.31	-0.50	0.36	-0.24	-0.17	-0.30	0.22	-0.10	1.00

The variance inflation analysis assesses the impact of multicollinearity of theses variables on the percent change in Southeast cotton acres (Table 3.9). Previous year cotton acres, the cotton to corn futures price ratio, previous year U.S. GDI, previous year ethanol price, and the cotton to soybean futures price ratio all exceed the 5% threshold.

Table 3.9: Southeast Percent Change Linear Regression Variance Inflation.

X Variable	VIF
Lag_Southeast_Acres	24.19
Cotton_Corn_ESR	4.895
Cotton_Wheat_ESR	2.319
Cotton_Change_FPDEC	4.056
Cotton_Soy_ESR	2.313
Cotton_Corn_FPRDEC	8.11
Lag_US_GDI	47.12
Lag_US_CPI	2.835
Lag_Ethanol	9.001
Cotton_Soy_FPRDEC	12.9
Lag_China_GDP	1.698

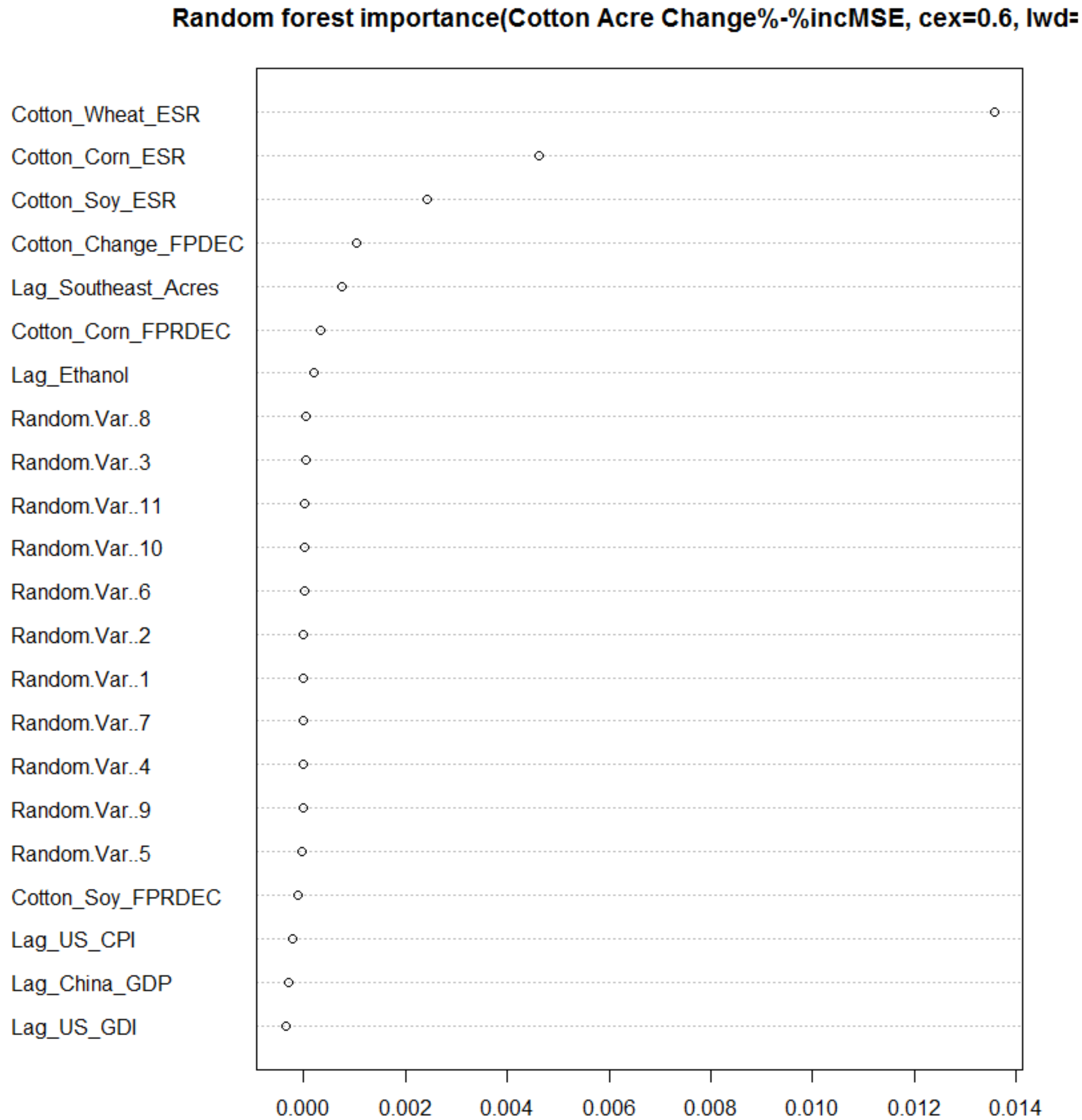
The Linear regression for the Delta had an R square was 0.61 and an adjusted R square was 0.39. The variable coefficients, standard errors, t statistic and P value are summarized below in Table 3.10. There are no variables that are significant.

Table 3.10: Summary of Linear Regression Independent Variables in Response to Percent Change in Southeast Cotton Acres.

Variable	Coefficients	Standard Error	t Stat	P-value
Intercept	-0.2105006	0.8002624	-0.2630	0.7955
Lag_Southeast_Acres	-0.0000003	0.0000002	-1.7112	0.1042
Cotton_Wheat_ESR	-0.0621768	0.0411341	-1.5116	0.1480
Lag_US_GDI	0.0000874	0.0000581	1.5054	0.1496
Cotton_Soy_ESR	-0.0061813	0.0045664	-1.3536	0.1926
Lag_Ethanol	-0.2662791	0.2066387	-1.2886	0.2138
Cotton_Soy_FPRDEC	4.6699963	6.7568539	0.6911	0.4983
Cotton_Change_FPDEC	-0.2003357	0.3447379	-0.5811	0.5684
Lag_China_GDP	0.0085800	0.0155938	0.5502	0.5889
Cotton_Corn_ESR	0.0180841	0.0475782	0.3801	0.7083
Cotton_Corn_FPRDEC	0.6926727	1.8860668	0.3673	0.7177
Lag_US_CPI	-0.0069619	0.0429480	-0.1621	0.8730

The Random Forest variable ranking shows several variables appearing before the first random variable (Figure 3.4). The variables in rank order are: cotton to wheat ending stock ratio, cotton to corn ending stock ratio, cotton to soybean ending stock ratio, change in cotton futures price, previous year Southeast cotton acres, the cotton to corn futures price ratio, and previous year ethanol price.

Figure 3.4: Southeast Region Random Forest Variable Ranking



3.2.4 Variable Assessment for the Southwest Region

The correlation matrix indicates a strong negative (<-0.50) relationships between the dependent variable, change in Southwest cotton acres and the previous year Southwest cotton acres (Table 3.11).

Table 3.11: Southwest Region Correlation Matrix

	Lag_Southwest_Acres	Cotton_Corn_ESR	Cotton_Wheat_ESR	Cotton_Change_FPDEC	Cotton_Soy_ESR	Cotton_Corn_FPRDEC	Lag_US_GDI	Lag_US_CPI	Lag_Ethanol	Cotton_Soy_FPRDEC	Lag_China_GDP	y
Lag_Southwest_Acres	1.00											
Cotton_Corn_ESR	0.26	1.00										
Cotton_Wheat_ESR	0.10	0.59	1.00									
Cotton_Change_FPDEC	-0.25	0.01	-0.16	1.00								
Cotton_Soy_ESR	-0.13	0.49	0.60	-0.11	1.00							
Cotton_Corn_FPRDEC	-0.17	-0.40	-0.46	0.15	-0.48	1.00						
Lag_US_GDI	0.28	0.35	0.34	0.19	0.36	-0.73	1.00					
Lag_US_CPI	0.13	-0.01	0.12	-0.40	0.00	0.14	-0.50	1.00				
Lag_Ethanol	0.17	0.37	0.42	-0.07	0.32	-0.79	0.63	0.07	1.00			
Cotton_Soy_FPRDEC	0.05	-0.25	-0.43	0.04	-0.50	0.88	-0.66	0.14	-0.76	1.00		
Lag_China_GDP	-0.08	0.30	0.32	0.06	0.14	-0.11	-0.03	-0.11	0.21	-0.17	1.00	
y	-0.60	-0.43	-0.46	0.42	-0.28	0.16	0.05	-0.34	-0.14	-0.01	-0.14	1.00

The variance inflation analysis assesses the impact of multicollinearity of theses variables on the percent change in Southeast cotton acres (Table 3.12). The cotton to corn futures price ratio and the cotton to soybean futures price ratio all exceed the 5% threshold. This would indicate that there is less impact caused by variance inflation due to multicollinearity in this region.

Table 3.12: Southwest Region Linear Regression Variance Inflation Analysis

X Variable	VIF
Lag_Southwest_Acres	2.23
Cotton_Corn_ESR	2.21
Cotton_Wheat_ESR	2.33
Cotton_Change_FPDEC	1.68
Cotton_Soy_ESR	2.40
Cotton_Corn_FPRDEC	8.28
Lag_US_GDI	7.94
Lag_US_CPI	3.51
Lag_Ethanol	4.91
Cotton_Soy_FPRDEC	7.11
Lag_China_GDP	1.75

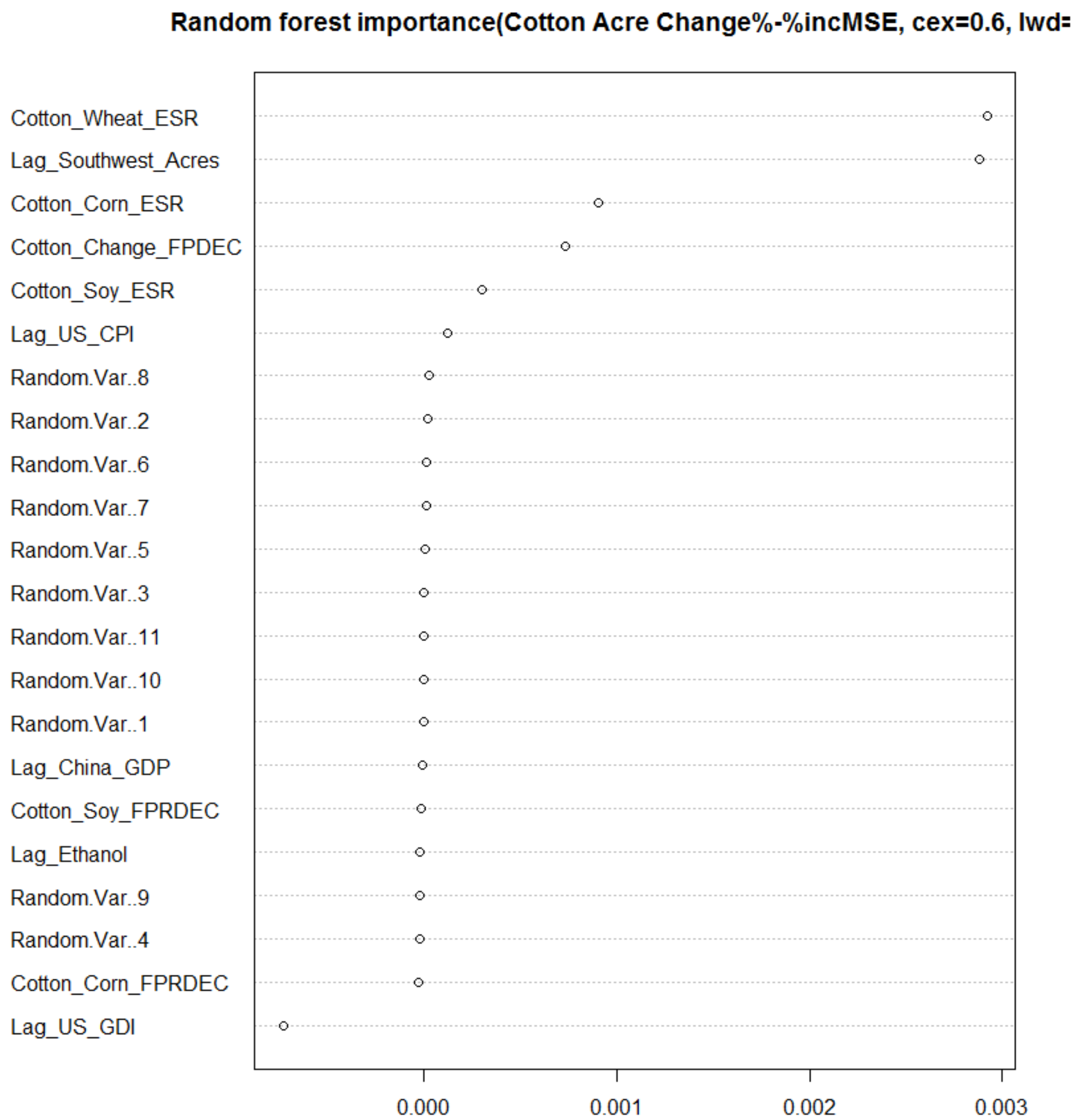
The Linear regression for the Delta had an R square was 0.78 and an adjusted R square was 0.64. The variable coefficients, standard errors, t statistic and P value are summarized below in Table 3.13. Previous year cotton acres and previous year U.S. GDI are significant at 0.01 level and the cotton to soybean ending stocks ratio is significant at the 0.05 level.

Table 313: Summary of Linear Regression Independent Variables in Response to Percent Change in Southwest Cotton Acres

Variables	Coefficients	Standard Error	t Stat	P-value
Intercept	0.95819250	0.31856989	3.0078	0.0076
Lag_Southwest_Acres	-0.0000002	0.00000003	-4.9494	0.0001**
Lag_US_GDI	0.00003610	0.00001216	2.9697	0.0082**
Cotton_Soy_ESR	-0.0056202	0.00237312	-2.3683	0.0293*
Cotton_Wheat_ESR	-0.0413370	0.02104079	-1.9646	0.0651
Lag_US_CPI	0.03441920	0.02437730	1.4119	0.1750
Lag_Ethanol	-0.1054794	0.07788302	-1.3543	0.1924
Lag_China_GDP	0.00463844	0.00805964	0.5755	0.5721
Cotton_Corn_FPRDEC	0.43357346	0.97155822	0.4463	0.6607
Cotton_Soy_FPRDEC	-1.0760370	2.55770383	-0.4207	0.6790
Cotton_Corn_ESR	-0.0028927	0.01629438	-0.1775	0.8611
Cotton_Change_FPDEC	0.00817118	0.11323569	0.0722	0.9433

The Random Forest variable ranking shows several variables appearing before the first random variable (Figure 3.5). The variables in rank order are: cotton to wheat ending stock ratio, previous year Southwest cotton acres, cotton to corn ending stock ratio, change in cotton futures price, cotton to soybean ending stock ratio, and previous year U.S. CPI.

Figure 3.5: Southwest Region Random Forest Variable Ranking



3.2.5 Variable Assessment for the West Region

The correlation matrix indicates a strong negative (<-0.50) relationships between the dependent variable, change in Southwest cotton acres and the cotton to wheat ending stocks ratio and the previous year U.S. CPI. A positive correlation (>0.50) exist with change in cotton price. (Table 3.14).

Table 3.14: West Region Correlation Matrix

	Lag_West_Acres	Cotton_Corn_ESR	Cotton_Wheat_ESR	Cotton_Change_FPDEC	Cotton_Soy_ESR	Cotton_Corn_FPRDEC	Lag_US_GDI	Lag_US_CPI	Lag_Ethanol	Cotton_Soy_FPRDEC	Lag_China_GDP	y
Lag_West_Acres	1.00											
Cotton_Corn_ESR	-0.20	1.00										
Cotton_Wheat_ESR	-0.22	0.59	1.00									
Cotton_Change_FPDEC	-0.31	0.01	-0.16	1.00								
Cotton_Soy_ESR	-0.32	0.49	0.60	-0.11	1.00							
Cotton_Corn_FPRDEC	0.67	-0.40	-0.46	0.15	-0.48	1.00						
Lag_US_GDI	-0.91	0.35	0.34	0.19	0.36	-0.73	1.00					
Lag_US_CPI	0.59	-0.01	0.12	-0.40	0.00	0.14	-0.50	1.00				
Lag_Ethanol	-0.61	0.37	0.42	-0.07	0.32	-0.79	0.63	0.07	1.00			
Cotton_Soy_FPRDEC	0.65	-0.25	-0.43	0.04	-0.50	0.88	-0.66	0.14	-0.76	1.00		
Lag_China_GDP	0.08	0.30	0.32	0.06	0.14	-0.11	-0.03	-0.11	0.21	-0.17	1.00	
y	-0.23	-0.40	-0.59	0.54	-0.40	0.24	-0.05	-0.50	-0.22	0.14	-0.11	1.00

The variance inflation analysis assesses the impact of multicollinearity of theses variables on the percent change in Southeast cotton acres (Table 3.15). The cotton to corn futures price ratio, previous year U.S GDI, previous year U.S. CPI, previous year Ethanol and the cotton to soybean futures price ratio all exceed the 5% threshold.

Table 3.15: West Region Linear Regression Variance Inflation Analysis

X Variable	VIF
Lag_West_Acres	16.51
Cotton_Corn_ESR	2.31
Cotton_Wheat_ESR	2.34
Cotton_Change_FPDEC	1.82
Cotton_Soy_ESR	2.23
Cotton_Corn_FPRDEC	8.10
Lag_US_GDI	8.11
Lag_US_CPI	5.23
Lag_Ethanol	7.16
Cotton_Soy_FPRDEC	6.50
Lag_China_GDP	2.33

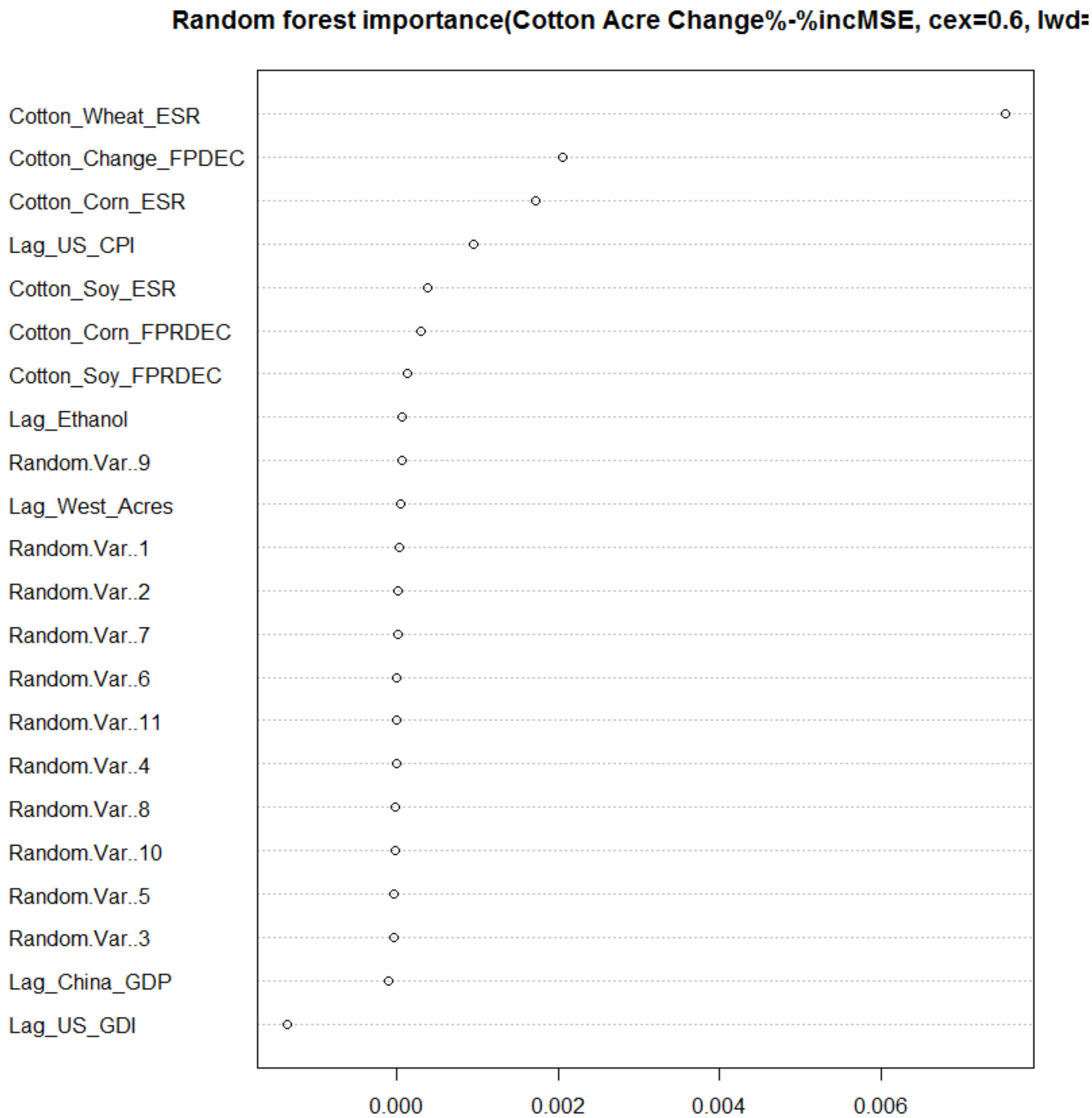
The Linear regression for the Delta had an R square was 0.78 and an adjusted R square was 0.64. The variable coefficients, standard errors, t statistic and P value are summarized below in Table 3.16. The previous year U.S. GDI is significant at 0.01 level and previous year cotton acres and the cotton to wheat ending stocks ratio are significant at the 0.05 level.

Table 3.16: Summary of Linear Regression Independent Variables in Response to Percent Change in West Cotton Acres.

Variables	Coefficients	Standard Error	t Stat	P-value
Intercept	1.6026373	0.4996185	3.2077	0.004
Lag_US_GDI	-0.0000569	0.0000167	-3.4041	0.0032**
Lag_West_Acres	-0.0000005	0.0000002	-2.6860	0.0151*
Cotton_Wheat_ESR	-0.0621895	0.0286897	-2.1677	0.0438*
Cotton_Change_FPDEC	0.2544248	0.1604721	1.5855	0.1303
Cotton_Soy_ESR	-0.0038805	0.0031084	-1.2484	0.2279
Lag_US_CPI	-0.0278872	0.0404958	-0.6886	0.4998
Cotton_Soy_FPRDEC	-1.6355768	3.3278500	-0.4915	0.6290
Lag_Ethanol	-0.0626023	0.1279016	-0.4895	0.6304
Lag_China_GDP	0.0049081	0.0126651	0.3875	0.7029
Cotton_Corn_ESR	0.0057021	0.0226882	0.2513	0.8044
Cotton_Corn_FPRDEC	0.1598237	1.3081846	0.1222	0.9041

The Random Forest variable ranking shows several variables appearing before the first random variable (Figure 3.6). The variables in rank order are: cotton to wheat ending stock ratio, change in cotton futures price, cotton to corn ending stock ratio, previous year U.S. CPI, cotton to soybean ending stock ratio, cotton to corn futures price ratio, cotton to soybean futures price ratio and previous year ethanol.

Figure 3.6: West Region Random Forest Variable Ranking



3.3 Model Development and Change in Acres Estimation

Multicollinearity between the exogenous variables was established in variable assessment through correlation between variables and as well in the variance inflation analysis. Estimates for change in cotton acres were calculated from two different modeling methods, linear regression and the combination of partial least squares/neural network (PLS/NN) and assessed for adjusted R Square. The Partial Least Squares is a technique was employed to reduce the effect of multicollinearity by maximising co-variance. (Yeniay and GÄokta 2002). The steps involved for this study were to scale the variables, create components from the PLS procedure and then predict or estimate the dependent variable y, using a neural network. The neural network was generated 100 times and the best performing model was selected for estimation (Qin and McAvoy 2000)

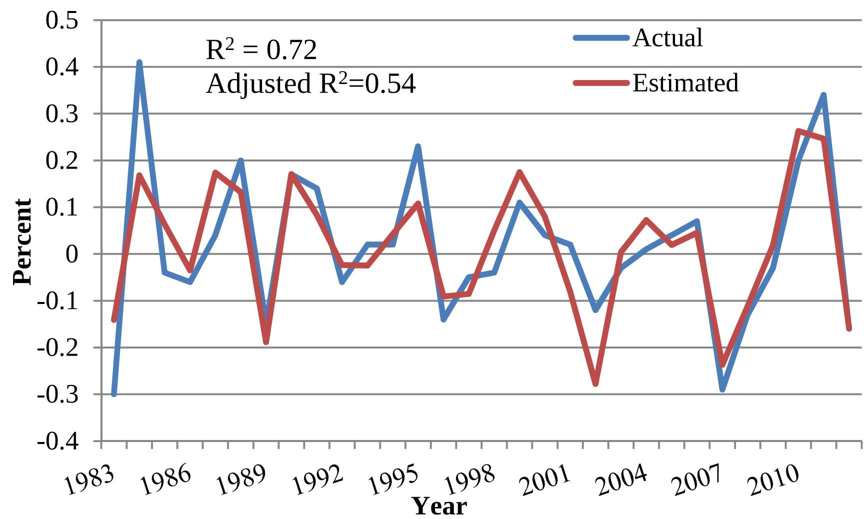
Both methods were used to estimate change in cotton acres by region and for the U.S. overall. The following subsections will cover the model results by region.

3.3.1 U.S. Model Results

The linear regression at the U.S. level had an adjusted R2 of 0.54 and did a reasonable job of estimating change in cotton acres (Figure 3.7). Based on the coefficient estimates described in Table 3.4, the model can be written as:

Change in US Cotton acres = 0.61+ 1.01 Cotton_Corn_ESR-0.03 Cotton_Wheat_ESR – 0.03 Cotton_Change_FPDEC-.007 Cotton_Soy_ESR+0.0000007 Lag_Cotton_ Acres _+ 0.89 Cotton_Corn_ FPRDEC+ 0.0005 Lag_US_ GDI_0.03 Lag_US_CPI-0.27Lag_Ethanol+0.64 Cotton_Soy_FPRDEC +0.008 Lag_China_GDP.

Figure 3.7: U.S. Percent Change in Cotton Acres Linear Regression Estimation



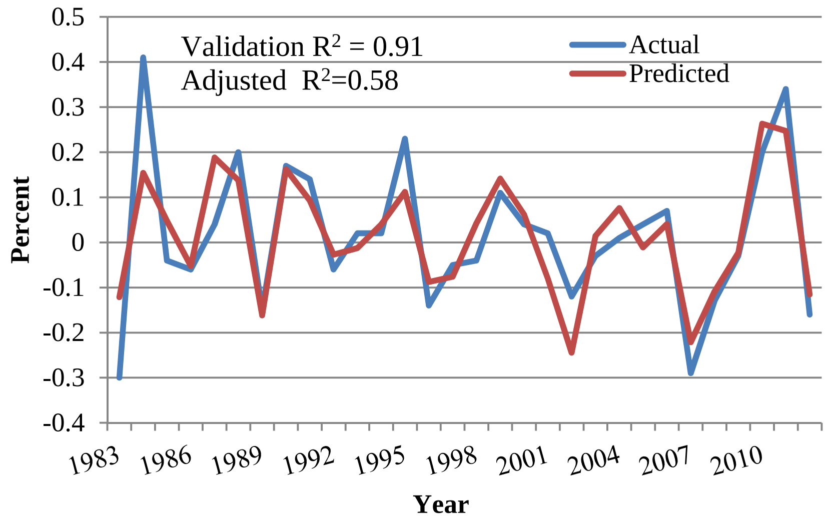
The PLS/NN model for the US change in cotton acres produced a 10-2-1 network with 25 weights, with linear output option and a decay set to 0.1. This means a weight is assigned to each linear combination between layers of nodes. The assigned weights are presented in table 3.17.

Table 3.17: US Change in Cotton Acres Neural Network Description and Weights

10-2-1 Network							
Pathway	Weight		Pathway	Weight		Pathway	Weight
b->h1	0.01		b->h2	-0.04		b->o	-0.05
i1->h1	-0.24		i1->h2	0.41		h1->o	-0.36
i2->h1	-0.12		i2->h2	0.21		h2->o	0.51
i3->h1	-0.09		i3->h2	0.13			
i4->h1	-0.13		i4->h2	0.16			
i5->h1	-0.08		i5->h2	0.11			
i6->h1	-0.1		i6->h2	0.16			
i7->h1	-0.08		i7->h2	0.06			
i8->h1	-0.05		i8->h2	0.03			
i9->h1	-0.03		i9->h2	0.02			
i10->h1	0.01		i10->h2	-0.01			

Estimates were generated using the PLS NN procedure. The adjusted R2 for this approach was 0.58, a slight improvement over the linear model. (Figure 3.8)

Figure 3.8: U.S. Percent Change in Cotton Acres PLS NN Estimation

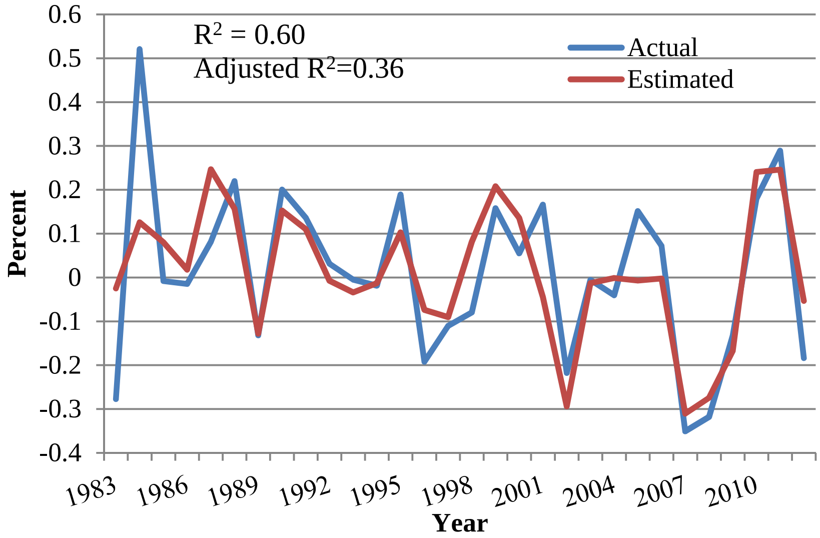


3.3.1 Delta Region Model Results

The Delta region consists of Arkansas, Mississippi, Missouri, Tennessee and Louisiana. The linear model adjusted R2 was 0.36 (Figure 3.9). Based on coefficients presented in Table 3.7, the model can be written as:

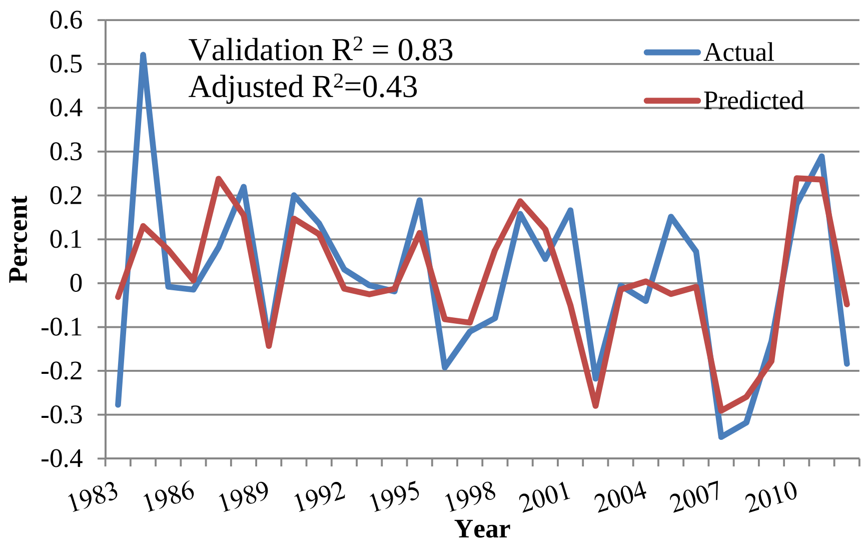
Change in Delta Cotton acres = 0.67- 0.0074 Cotton_Soy_ESR+0.0000901 Lag Delta Acres-0.2Lag_Ethanol+0.00002Lag_US_GDI+0.76Cotton _Corn_FPRDEC-0.01Cotton_Wheat_ESR-0.01Cotton to corn-1.04 Cotton_Soy_FPRDEC +0.0012Lag_China_GDP+.003Lag_US_GDI+0.01Cotton_Change_FPRDEC.

Figure 3.9: Delta Change in Cotton Acres Linear Regression Estimation



The PLS NN improved adjusted R2 to 0.43 and is closer in the years 1986, 1995, 1999, 2000 and 2002. The PLS NN was not as close as the linear model in 2005, 2007 and 2008 (Figure 3.10).

Figure 3.10: Delta Change in Cotton Acres PLS NN Estimation



The PLS/NN model for the Delta change in cotton acres produced a 9-1-1 network with 12 weights, with linear output option and a decay set to 0.2. This means a weight is assigned to each linear combination between layers of nodes. The assigned weights are presented in Table 3.18.

Table 3.18: Delta Change in Cotton Acres Neural Network Description and Weights

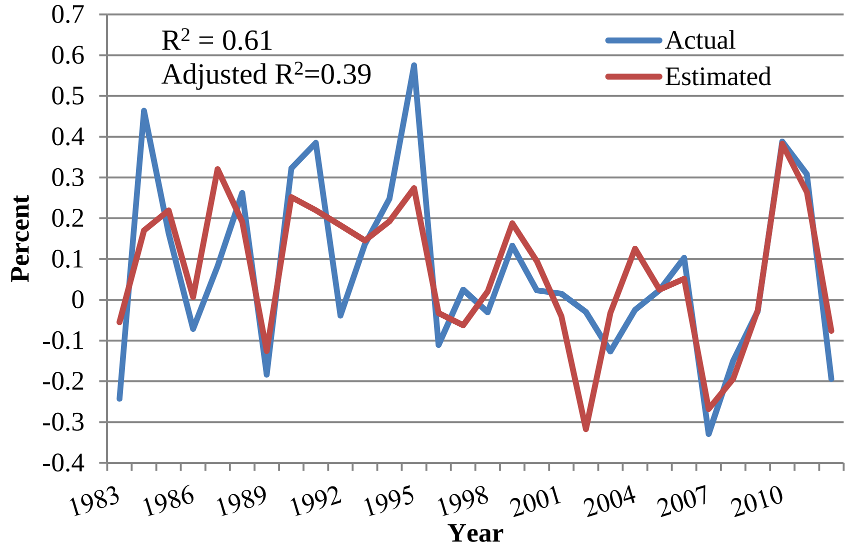
9-1-1 Network				
Pathway	Weight		Pathway	Weight
b->h1	-0.09		b->o	-0.05
i1->h1	-0.51		h1->o	-0.36
i2->h1	-0.25			
i3->h1	-0.11			
i4->h1	-0.13			
i5->h1	-0.01			
i6->h1	-0.07			
i7->h1	-0.01			
i8->h1	-0.04			
i9->h1	0			

3.3.3 Southeast Region Model Results

The Southeast region was characterized as Georgia, North Carolina, South Carolina, and Alabama. Like the Delta, the models for the Southeast region didn't perform as well as the overall U.S. model. Linear regression for the Southeast Region has an adjusted R^2 of 0.39 (Figure 3.11). Based on the linear coefficients in Table 3.10, the linear model can be written as:

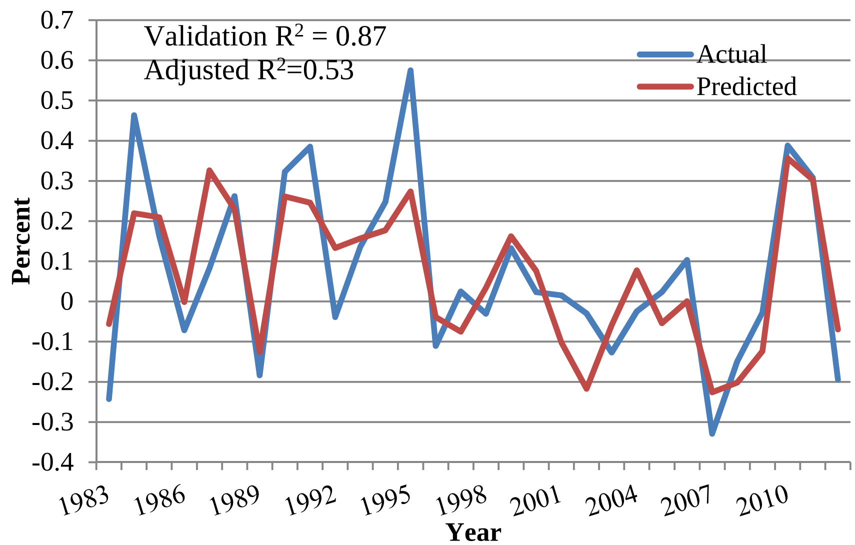
Change in Southeast Cotton acres =- 021 -0.0000003 Lag_Southeast_Acres -0.06 Cotton_Wheat_ESR +0.00009 Lag_US_GDI -0.006 Cotton_Soy_ESR- 0.26 Lag_Ethanol + 4.67 Cotton_Soy_FPRDEC- 0.2Cotton_Change_FPDEC+ 0.008 Lag_China_ GDP +0.02 Cotton_Corn_ESR+0.69 Cotton _Corn_FPRDEC- 0.01Cotton_Wheat_ESR-0.01Cotton to corn-1.04 Cotton_Soy_FPRDEC - 0.006Lag_US_GDP.

Figure 3.11: Southeast Region Change in Cotton Acres Linear Regression Estimation



Estimation does improve with the PLS/NN approach and is noticeably closer in 2002. The adjusted R^2 for this approach is 0.53 (Figure 3.12). Neither the linear model or the PLS/NN approach caught the upward surges in 1984 and 1995.

Figure 3.12: Southeast Percent Change in Cotton Acres PLS NN Estimation



The PLS/NN model for the Southeast change in cotton acres produced a 9-2-1 network with 23 weights, with linear output option and a decay set to 0.1. The assigned weights are presented in Table 3.19.

Table 3.19: Southeast Change in Cotton Acres Neural Network Description and Weights

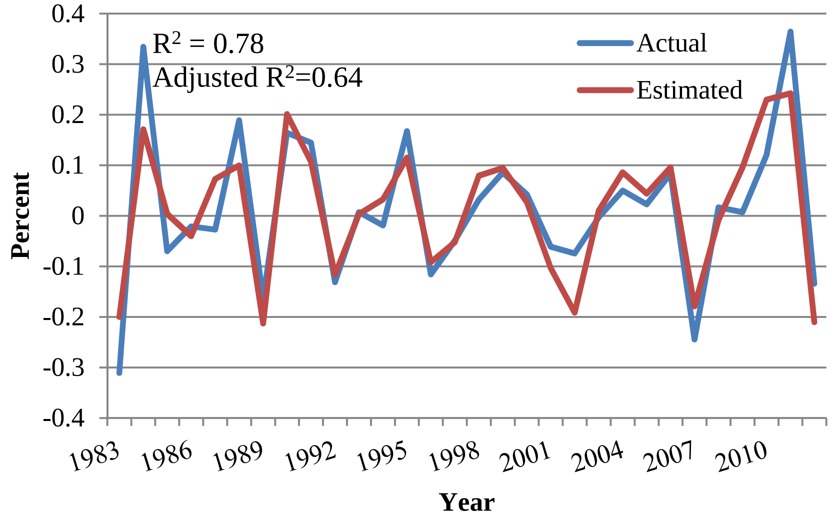
9-2-1 Network							
Pathway	Weight		Pathway	Weight		Pathway	Weight
b->h1	0.28		b->h2	-0.04		b->o	0.24
i1->h1	-0.7		i1->h2	0.41		h1->o	-0.64
i2->h1	-0.44		i2->h2	0.21		h2->o	0.33
i3->h1	-0.12		i3->h2	0.13			
i4->h1	-0.14		i4->h2	0.16			
i5->h1	-0.03		i5->h2	0.11			
i6->h1	-0.21		i6->h2	0.16			
i7->h1	-0.03		i7->h2	0.06			
i8->h1	-0.18		i8->h2	0.03			
i9->h1	0.03		i9->h2	0.02			

3.3.4 Southwest Region Model Results

The Southwest region included Texas, Oklahoma and Kansas. The linear model performed well for this region with an adjusted R^2 of 0.64 (Figure 3.13). Note that unlike the Delta and Southeast, the increase change increase in 1984 was somewhat accounted for in this model. Table 3.13 provides the linear coefficients so the linear model can be written as:

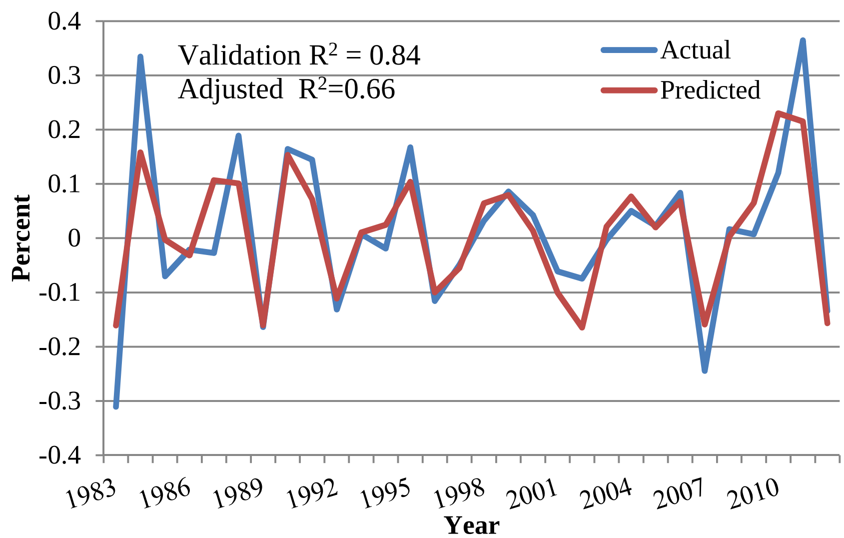
Change in Southwest Cotton Acres = 0.95 -1.72E-07 Lag_Southwest_Acres + 3.61E-05 Lag_US_GDI -0.005 Cotton_Soy_ESR -0.0413370198331225 Cotton_Wheat_ESR +0.034 Lag_US_CPI -0.105479380442622 Lag_Ethanol + 0.005 Lag_China_GDP +0.43 Cotton_Corn_FPRDEC -1.08 Cotton_Soy_FPRDEC -0.003Cotton_Corn_ESR +0.008 Cotton_Change_FPDEC.

Figure 3.13: Southwest Region Change in Cotton Acres Linear Regression Estimation



The PLS NN demonstrated only a slight improvement in performance with an adjusted R^2 of 0.66. The estimates were closer to actual than those for the linear model in 1989, 1993, 1997, 1999, and 2005 (Figure 3.14).

Figure 3.14: Southwest Region Percent Change in Cotton Acres PLS NN Estimation



The PLS/NN model for the Southeast change in cotton acres produced a 9-1-1 network with12 weights, with linear output option and a decay set to 0.1. The assigned weights are presented in Table 3.20.

Table 3.20: Southwest Change in Cotton Acres Neural Network Description and Weights

9-1-1 Network				
Pathway	Weight		Pathway	Weight
b->h1	-0.06		b->o	-0.29
i1->h1	0.47		h1->o	0.62
i2->h1	0.21			
i3->h1	0.15			
i4->h1	0.24			
i5->h1	0.17			
i6->h1	0.08			
i7->h1	0.06			
i8->h1	-0.02			
i9->h1	0.01			

3.3.4 West Region Model Results

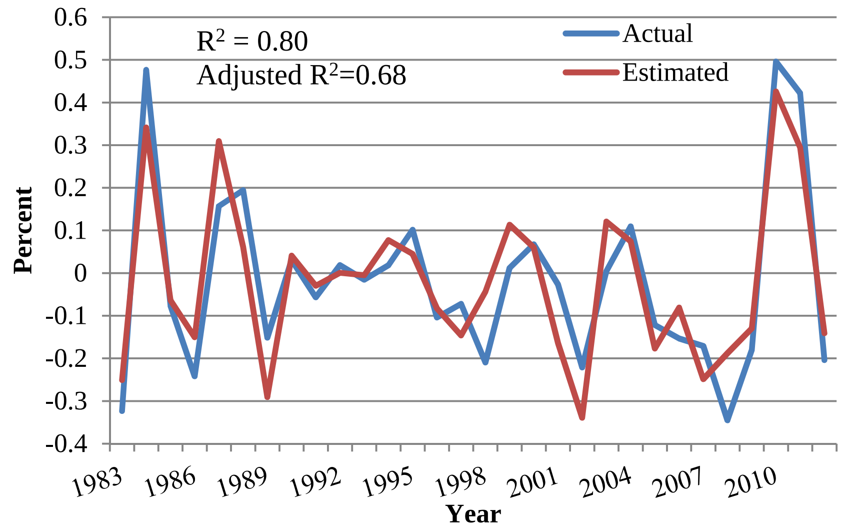
The West region included California, New Mexico and Arizona. Analysis of change in acres in this region yielded results more similar to the Southwest. The linear regression model for this region performed reasonably well with an adjusted R^2 of 0.68. The estimates from this model are more closely capturing the increase in percent change in upswing years than in any of the other regions (Figure 3.15). The linear model may be written as (Table 3.16):

Change in West Cotton Acres = 1.6 -5.69E-05 Lag_US_GDI -5.44E-07 Lag_West_Acres

-0.06 Cotton_Wheat_ESR + 0.25 Cotton_Change_FPDEC -0.004Cotton_Soy_ESR

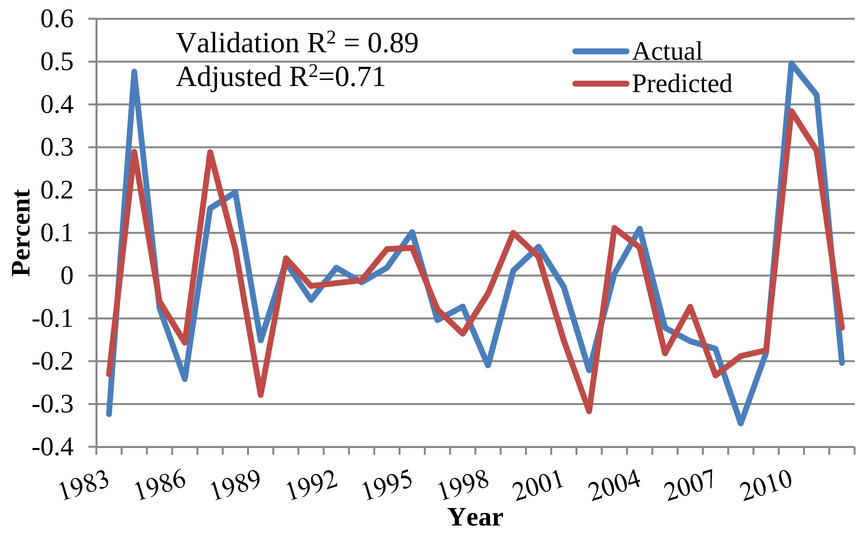
-0.03 Lag_US_CPI -1.64 Cotton_Soy_FPRDEC -0.06 Lag_Ethanol + 0.005 Lag_China_GDP + 0.006 Cotton_Corn_ESR + 0.169 Cotton_Corn_FPRDEC.

Figure 3.15: West Region Percent Change in Cotton Acres Linear Regression Estimation



The PLS NN provides a slight improvement to adjusted R Square but the estimates from the linear model were closer to the actual change in acres in 1984 and 1998 (3.17)

Figure 3.16: West Region Percent Change in Cotton Acres, PLS NN Estimation



The PLS/NN model for the Southeast change in cotton acres produced a 9-3-1 network with 34 weights, with linear output option and a decay set to 0.1. The assigned weights are presented in Table 3.21.

Table 3.21: West Change in Cotton Acres Neural Network Description and Weights

9-3-1 Network										
Pathway	Weight		Pathway	Weight		Pathway	Weight		Pathway	Weight
b->h1	0.05		b->h2	-0.2		b->h3	0.05		b->o	0.02
i1->h1	-0.2		i1->h2	0.43		i1->h3	-0.2		h1->o	-0.34
i2->h1	-0.09		i2->h2	0.1		i2->h3	-0.09		h2->o	0.67
i3->h1	-0.13		i3->h2	0.28		i3->h3	-0.13		h3->o	-0.34
i4->h1	-0.14		i4->h2	0.2		i4->h3	-0.14			
i5->h1	-0.09		i5->h2	0.13		i5->h3	-0.09			
i6->h1	-0.07		i6->h2	0.14		i6->h3	-0.07			
i7->h1	-0.03		i7->h2	0.14		i7->h3	-0.03			
i8->h1	-0.05		i8->h2	0.1		i8->h3	-0.05			
i9->h1	-0.03		i9->h2	0.05		i9->h3	-0.03			

CHAPTER IV SUMMARY AND CONCLUSIONS

The utility of this study targets estimated cotton acres in a 12-24 month time horizon suitable for decision support for seed production. An exploration of supply and demand factors help to identify regional differences in response to these stimuli.

The Delta region had higher variance inflation indicating strong effect of multicollinearity.

While no variables were significant in the regression, the influence of alternative crops expresses strongly in this region (Table 4.1.) All corn related variables (ending stocks, futures price and ethanol price) were influential according to the Random Forest variable ranking. The change in acres are harder to estimate than in the Southwest and West; the PLS/NN provided improved model performance increasing the adjusted R Square by seven percent.. The Delta and Southeast regions respond to the same variables, though cotton to corn ending stock ratio was stronger than wheat.

Table 4.1: Variable Importance in Linear Regression and Random Forest for the Delta and the Southeast

	Delta		Southeast	
Variable	Linear Regression	Random Forest	Linear Regression	Random Forest
Lag_Acres		X		X
Cotton_Corn_ESR		X		X
Cotton_Wheat_ESR		X		X
Cotton_Change_FPDEC		X		X
Cotton_Soy_ ESR		X		X
Cotton_Corn_ FPRDEC		X		X
Lag_Ethanol		X		X

The Southeast region responds similar to the Delta and the same variables were important for both regions (Table 4.1). Like the Delta, the Southeast has high variance inflation as well. There were no significant variables from the linear regression model from this region either. The Random Forest did highlight the importance of alternative crops; especially cotton to wheat ending stocks ratio. Relative to the Delta, cotton to wheat ending stocks ratio is more highly ranked than cotton to corn. Change in acres are harder to estimate than in the Southwest and West, again the PLS/NN provided improved model performance with 14% improvement in adjusted R Square.

The Southwest represents the largest production area and it also has fewer alternative crops. The variance inflation is lower in this region and the linear model performs well. Lag cotton acres, change in cotton price, lag U.S. GDI and the cotton to soybean ending stock ratio are important according to the linear regression model (Table 4.2). The cotton to wheat ending stock ratio is very close to significance. The cotton to wheat ending stocks ratio ranked highly according to the Random Forest, an observation that is reinforced by the negative correlation (-0.46) between the change in acres and this variable. The lag US CPI was important in the Random Forest while lag US GDP was the lowest ranked, this may be because of the correlation between this two variables. The influence of these macroeconomic variables may indicate sensitivity to consumption growth.

Table 4.2: Variable Importance in Linear Regression and Random Forest for the Southwest

	Southwest	
Variable	Linear Regression	Random Forest
Lag_Acres	X	X
Cotton_Corn_ESR		X
Cotton_Wheat_ESR		X
Cotton_Change_FPDEC	X	X
Cotton_Soy_ ESR		X
Cotton_Corn_ FPRDEC		X
Lag_US_GDI	X	
Lag_US_CPI		X

The West is the smallest growing region and behaves most similarly to the Southwest. The linear model works well for estimating change in acres for this region as well. The linear regression model identified lag U.S. GDI, lag acres, and the cotton to wheat ending stocks ratio as significant (Table 4.3). The cotton to wheat ending stock ratio was the highest rank in the Random Forest. Change in cotton price, cotton to corn ending stock ratio, lag U.S. CPI, cotton to soybean ending stock ratio, cotton to corn futures price ratio as well as cotton to soybean futures price ratio and lag ethanol price were also identified as important in the Random Forest. This is the only region that demonstrated response with the cotton to soybean futures price ratio, this may be because of the correlation with macroeconomic variables.

Table 4.3: Variable Importance in Linear Regression and Random Forest for the West

	West	
Variable	Linear Regression	Random Forest
Lag_Acres	X	
Cotton_Corn_ESR		X
Cotton_Wheat_ESR	X	X
Cotton_Change_FPDEC		X
Cotton_Soy_ ESR		X
Cotton_Corn_ FPRDEC		X
Lag_US_GDI	X	
Lag_US_CPI		X
Lag_Ethanol		X
Cotton_Soy_ FPRDEC		X

Across regions previous year acres were the variable most often identified as important. This is consistent with a reluctance to switch as noted by Fanin and Paxton (2011). The cotton to wheat ending stock relationship was also very strong due to availability of this crop as most likely alternative to cotton in the West and Southwest. The cotton to corn ending stock ratio, the change in cotton price and the futures price ratio between cotton and corn were also important in all regions.

The cotton to soybean ending stock ratio was important across regions as well, even in regions where soybean is not a good alternative, such as the Southwest. This could be an artifact of the ratio with cotton and cotton's relative strength or decline regardless of the crop or it could be a result of the moderate correlation with previous year U.S. GDI. Previous year China GDP was not important in any region.

The culmination of this work is to inform seed production decisions with regional acreage estimates 12- 24 months ahead of sales.   This set of exploratory analyses prepares a foundation for taking additional steps in variable selection.

Next steps for variable selection would be to improve variable selection, test and extend the time horizon, and test other estimation approaches. Variable selection could be improved by 1) adding region specific variables such as futures/contract and ending stocks for crops such as sorghum in Texas. Wheat futures were not explored but may also provide additional predictive performance 2) the refinement and addition of consumption variables and 3) addressing the impact of China more directly.

Extending the time horizon for estimation and prediction involves assessing variables have for availability and relevance in the appropriate time horizon as with the futures price ratios or the ability to forecast the variables themselves as with U.S. CPI. Other modeling approaches can be explored as well. Relationships between the futures price, ending stocks and consumption change. Multiple decision time points exist and the information is needed multiple times up during the manufacturing cycle, thus a Bayesian approach is worth investigating

The model development work in the study shows that it will be important to address multicollinearity and that for the regions with more alternative crop options, a non-linear approach may improve model performance.

REFERENCES

Baffes, J., and G. Gohou. "The Co-movement between Cotton." *World Bank eLibrary.* March 2005. http://elibrary.worldbank.org/docserver/download/3534.pdf?expires=1355496508&id=id&accname=guest&checksum=83297CC6183111E8D50331ED945E1EE3 (accessed December 9, 2012).

Fannin, James Matthew, and Kenneth W. Paxton. "Measuring the Relative Profitability of Mid-South Cotton Production from an Alternative Gin Seed Rebate Model." *2011 Annual Meeting.* Corpus Christi, TX: Southern Agricultural Economics Association, 2011.

Foreman, Linda F. *Characteristics and Production Costs of U.S. Cotton Farms, 2007; EIB-104.* December 2012. http://purl.umn.edu/142358 .

GATT. "Agreement Regarding International Trade in Textiles." *GATT Digital Library:1947-1994.* December 20, 1973. http://sul-derivatives.stanford.edu/derivative?CSNID=90890198&mediaType=application/pdf (accessed December 12, 2012).

Hudson, Darren, Suwen Pan, Maria Erlinda M. Mutuc, Samantha Yates, and Don E. Ethridge. *Global Cotton Baseline 2010/11 - 2020/21.* Austin, TX: Texas Tech University, Department of Agricultural and Applied Economics, 2011.

ICTSD. "Cotton: What could a Doha deal mean for trade?" *International Centre for Trade and Sustainable Development.* November 2010. http://ictsd.org/downloads/2011/12/cotton-what-could-a-doha-deal-mean-for-trade.pdf (accessed December 09, 2012).

Isengildina-Massa, Olga, and Stephen MacDonald. *U.S. Cotton Prices and the World Cotton Market; Forecasting and Structural Change; Economic Research Report.* Washington, D.C.: USDA-ERS, 2009.

Liang, Yan, J. Corey Miller, Ardian Harri, and Keith H. Coble. "Crop Supply Response under Risk." May 2011. http://purl.umn.edu/104615.

MacDonald, S, and T. Vollrath. "The Forces Shaping World Cotton Consumption After the Multi-fiber Arrangement (CWS-05c-01)." *USDA, Electronic Outlook Report from the Economic Research Service.* April 2005. http://www.ers.usda.gov (accessed November 20, 2012).

Monsanto. "History of Bollgard Cotton." *Monsanto.com.* December 7, 2012. http://www.monsanto.com/improvingagriculture/Pages/history-of-bollgard-cotton.aspx (accessed December 7, 2012).

—. "SEC Filings, 10K reports 2006-2012." *Monsanto Investor Relations.* 2012. http://www.monsanto.com/investors/Pages/sec-filings.aspx (accessed December 9, 2012).

Mutuc, Maria Erlinda M., and Darren Hudson. "DECLINING COTTON ACREAGE IMPACTS ON U.S. COTTON GINNING INDUSTRY STRUCTURE AND COSTS." *2010 Cotton Beltwide Conferences, New Orleans, Louisiana, January, 2010.* 01 2010. http://purl.umn.edu/96673 (accessed 01 2013).

Neves, Marcos Fava, Mairun Junqueira Alves Pinto, and José Carlos de Lima Júnior. "Mapping and Quantification of the Cotton Chain in Brazil." *International Food and Agribusiness Management Review, Volume 15, Issue 4, 2012*, 2012: 189-193.

Pan, Suwen, Darren Hudson, and Don E. Ethridge. "Market Structure Impacts on Market Distortions from Domestic Subsidies: The U. S. Cotton Case." *Estey Centre Journal of International Law and Trade Policy; Volume 11, Number 2, Summer 2010*, 2010: 417-435.

Parrott, Scott D., and Christopher S. McIntosh. "NONCONSTANT PRICE EXPECTATIONS AND ACREAGE RESPONSE: THE CASE OF COTTON PRODUCTION IN GEORGIA." 1996. http://purl.umn.edu/15235.

Qin, S.J., and T.J. McAvoy. "Nonlinear PLS modeling using neural networks." *Computers & Chemical Engineering* , 2000: DOI:10.1016/0098-1354(92)80055-E.

Strobl, C., A Boulesteix, A. Zeileis, and T. Hothorn. *Bias in random forest variable importance measures: Illustrations, sources and a solution.* January 25, 2007. http://www.biomedcentral.com/1471-2105/8/25 (accessed January 12, 2013).

Tribble, Camille M, Christopher S McIntosh, and Michael E. Wetzstein. "GEORGIA COTTON ACREAGE RESPONSE TO THE BOLL WEEVIL ERADICATION PROGRAM." *Journal of Agricultural and Applied Economics>Volume 31, Number 03, December 1999*, 1999: 499-5-6.

Tronstad, R., and R. Bool. "U.S. Cotton Acreage Response Due to Subsidized Crop Insurance." *Agricultural & Applied EconomicsAssociation 2010 AAEA, CAES, & WAEA Joint Annual Meeting, Denver, Colorado.* July 25-27, 2010. http://purl.umn.edu/61455 (accessed January 8, 2013).

USDA- NASS. *Data Query.* December 1, 2012. http://quickstats.nass.usda.gov/results/D216BBF1-1737-33A8-8A30-36341B92A8E3?pivot=short_desc (accessed December 1, 2012).

USDA-NASS. *Usual Planting and Harvest Report 2010.* Washington, District of Columbia, October 2010.

Wikipedia. "Multi-fiber Agreement." *Wikipedia.* October 22, 2011. http://en.wikipedia.org/wiki/Multi_Fibre_Arrangement (accessed December 12, 2012).

—. *Wikipedia:Random forest.* March 14, 2013. http://en.wikipedia.org/wiki/Random_forest (accessed March 15, 2013).

Yearman, K., Gluckman, A. "Falling Off a Cliff." *Dollars and Sense; Real World Economics.* September 05, 2005. http://www.dollarsandsense.org/archives/2005/0905yearman.html (accessed December 12, 2012).

Yeniay, OzgÄur, and Atilla GÄokta. "A Comparison of Least Squares Regression with Other Methods of Prediction." *Hacettepe Journal of Mathematics and Statistics,* 2002: Volume 31, 99-111.