2017

Factors Influencing Illinois Soybean Yield Increase

Sydney Bollinger
sydneybollinger@siu.edu

FACTORS INFLUENCING ILLINOIS SOYBEAN YIELD INCREASE

by

Sydney Bollinger

B.S., Southern Illinois University Carbondale, May 2017

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

Department of Agribusiness Economics
in the Graduate School
Southern Illinois University Carbondale
December, 2017

RESEARCH PAPER APPROVAL

FACTORS INFLUENCING ILLINOIS SOYBEAN YIELD INCREASE

By

Sydney Bollinger

A Research Paper Submitted in Partial

Fulfillment of the Requirements

For the Degree of

Master of Science

in the field of Agribusiness Economics

Approved by:

Dr. Dwight R. Sanders

Graduate School
Southern Illinois University Carbondale
November 5, 2017

AN ABSTRACT OF THE RESEARCH PAPER OF

SYDNEY BOLLINGER, for the Master of Science degree in AGRIBUSINESS ECONOMICS, presented on NOVEMBER 5th, 2016 at Southern Illinois University Carbondale.

TITLE: FACTORS INFLUENCING ILLINOIS SOYBEAN YIELD INCREASE

MAJOR PROFESSOR: Dr. Dwight R. Sanders

Soybean productivity is among many important topics currently in Illinois agriculture. With corn prices being so low the last few years, many people are relying on their soybean crop to produce very well. Also, with the increasing disapproval of genetically modified crops, it is important to recognize if these crops making a difference in production. It is obvious that weather plays a huge role in crop productivity. This research is intended to identify other specific variables that influence soybean yield in a given year. Much research has been done in the past identifying the weather variables that are most important in relation to bushels per acre harvested, but little research has truly delved into the technological advances of the last 50 years. Using multiple regression analysis, this paper pinpoints the specific sources behind soybean yield increase that include July precipitation, August precipitation, and August temperature. Conversely, the regression analysis surprisingly did not find any significance in double cropping, biotech acres planted or row width used.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The soybean was first introduced to the United States in 1765 when Samuel Bowen requested that Henry Yonge plant them on his farm. Mr. Bowen brought the soybeans from China to plant in Thunderbolt, Georgia. At this time, the soybeans were used to manufacture soy sauce and vermicelli to export to England. Bowen died shortly after in December of 1777, and his soybean enterprise in Georgia terminated (Hymowitz, 1987).

The second time soybeans were introduced to North America was in 1770 by Benjamin Franklin. He had seeds sent from London to Philadelphia where the botanist John Bartram received them. There, Dr. James Mease came to the conclusion that the seeds grew well in Pennsylvania and should be cultivated. It wasn't until 1851 that soybeans were brought to Illinois by Dr. Benjamin Franklin Edwards. He obtained seeds from the Japanese in San Francisco and brought them back to Alton, Illinois. There, he gave the seeds to Mr. John H. Lea who planted them in his garden in the summer of 1851. After harvest, he circulated them to several people including Mr. J. J. Jackson of Davenport, Iowa, who was the first person to grow soybeans in Iowa. By the end of 1854, soybeans were thoroughly dispersed and evaluated by farmers throughout the United States (Hymowitz, 1987).

This introduction by Dr. Benjamin Franklin Edwards took place approximately 3 years prior to the much better known introduction of soybeans from Japan by the Perry Expedition. Commodore Matthew Perry opened up Japan to western trade and the expedition's surgeon, Dr. Daniel Green, witnessed an unusual bean growing called the Japan Pea. The expedition's agriculturist, Dr. James Morrow then obtained the seeds and sent them to be distributed. Because these events happened so closely together, and the previous introduction was not properly documented, it is difficult to know where a particular farmer's soybeans originated from after this time period (Hymowitz, 1987).

Production of soybeans is expanding rapidly worldwide due to increasing meat consumption. Internationally, about 75% of soy is used for animal feed with the majority of that going to poultry and swine. 19% is used and make soy oil and a mere 6% of soy produced worldwide is directly used as food. This is mainly in Asian countries where whole beans may be eaten as a vegetable or included into tofu or soy sauce. 2% of that 6% is additionally processed into flours or protein additives. Lastly, around 2% of total soy production is used to produce biodiesel (Soy and it's Uses, 2017).

Agriculture in Illinois directly and indirectly creates about 1.5 million jobs. Soybean acres cover approximately nine million acres accounting for 25% of Illinois' total acreage. From that, an average of 420 million bushels are produced making Illinois one of the top producers in the country. Soybeans make a huge part of the state's agricultural output being worth nearly $6 billion per year in direct sales. Roughly half of the crop is exported with the other half being crushed (Illinois Soybeans Matter, 2017).

The main objective of this paper is to identify the variables that significantly impact Illinois soybean yields. Farmers and researchers across the world are constantly trying to improve their yields so it is important to recognize the aspects of their operations that are influencing yield the most. A multiple regression model will examine the effects of precipitation, temperature, row spacing, genetically modified crops, and double cropping from 1974 – 2016. This regression will simultaneously demonstrate the significance or insignificance of each independent variable.

This information will be valuable to farmers and researchers focusing on increasing yield, especially considering much of the previous research done on this topic was limited to weather variables. Also, with the growing disapproval of genetically modified organisms, it is important to see if these factors have had a major impact on yield.

CHAPTER 2

REVIEW OF LITERATURE

In this research, the ordinary least squares (OLS) linear multiple regression model will be used to determine significant variables. “Regression analysis is concerned with the study of the dependence of one variable, the dependent variable, on one or more other variables, the explanatory variables, with a view to estimating and/or predicting the (population) mean or average value of the former in terms of the known or fixed (in repeating sampling) values of the latter” (Gujarati, 2009 p. 15). Since this research is studying “the dependence of one variable on more than one explanatory variable” (Gujarati, 2009 p. 21), it is known as multiple regression analysis. Francis Galton first introduced the term regression in a study relating height of parents to their children. Height of a child tended to “regress” towards the average height of the population as a whole. His friend, Karl Pearson collected more than one thousand records of families heights and then found the same “regressing” of tall fathers to have sons shorter than them and short fathers to have sons taller than them (Gujarati, 2009 p. 15).

The first time glyphosate resistant soybeans (roundup ready crops) were tested in the field was in 1989 and they sold in the market in 1996 for the first time. This was huge for soybean farmers and the adoption of the technology was extremely rapid. In the beginning, many growers found they actually tended to yield less than conventional varieties. To be specific, it was around 5 to 7% less. It was never discovered why exactly this yield drag happened, but it is important to grasp that there was never a yield increase promise with the introduction of GM (genetically modified) crops. There was only a promise for life to get a little easier in the weed control category (Green, 2009).

A previous study done by David Goldblum delves into the effect of temperature and precipitation and soybean yield in Illinois. He looks into this in the study of the effect of climate change on our agricultural yields. A correlation analysis was done with crop yield compared to monthly precipitation as well as average monthly temperature data. It was discovered that soybean yield is more sensitive to climate than corn yield. (Goldblum 2009).

Tannura, Irwin, and Good (2008) conducted a study similar to Goldblum looking into the weather variables impacting yield in the United States Corn Belt. This included Illinois, Indiana, and Iowa because they have “similar weather and crop development timescales and they represent nearly half of United States corn and soybean production” (Tannura, 2008 pg. 38). They observed monthly mean temperatures for May through September and monthly precipitations for May through September as well. This analysis concluded “yields were reduced by unfavorable weather by a much larger amount than they were increased by favorable weather.” Moreover, it was found that soybean yields were most affected by technology, precipitation in June, precipitation in July, and especially precipitation in August. July and August temperatures also showed some significance (Tannura 2008). The purpose of this research is to test these findings on a different scale with the data chosen.

CHAPTER 3

DATA

The data used in this research relates to totals from soybeans raised in Illinois. Average yield is my dependent variable. This can be displayed in figure 1 of the appendix. This information was collected from the United States Department of Agriculture's National Agriculture Statistics Service QuickStats program from the year 1974 to 2016 for a total of 43 years. Yield is measured as bushels per harvested acre. The reason for collecting these specific years is due to the availability of other variables such as percent of biotech acres, width of rows, and percentage of acres double cropped. These independent variables were also collected from NASS QuickStats. These variables were chosen to represent the major changes in technology and improvements in farming techniques over the last 43 years. A trend in higher yields in more recent dates, higher percentages of GMOs in more recent dates, and smaller row widths in more recent dates were detected. However, double cropping does not have a visible trend. This can be seen in figures 2, 3, 4 and 5 of the appendix.

In looking for data on these variables, some missing observatioins were in the records of NASS QuickStats. The first four collected years of percentage of acres double cropped were missing from the database. However, the national percentages were available. Given that Illinois double crop acres are generally 1% higher than the national average, years 1974 - 1977 were estimated using that knowledge. Similarly, years 1996 – 1999 in the category percentage of acres biotech were missing from the database. These were estimated using the understanding that the United States national average was about 1% more than Illinois.

Aside from the yield variables, the rest of the data used was collected from the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental

Information. Precipitation for the months of June, July, August, and September from 1974 to 2016 and average monthly temperature for the months of June, July, August, and September from 1974 to 2016 were recorded. The precipitation is simply the cumulative daily rainfall for the specified month measured in inches. The monthly temperature is identified as the daily average temperature measured in degrees Fahrenheight during each month. These four months were chosen because they are the months where the precipitation and temperature were proven to be most important to the production of the soybean plant (Sanders, Altman, Ferraro 2013).

CHAPTER 4

METHODS

As previously stated, the regression used in this research is an ordinary least squares multiple linear regression. This is used to show the relationship between variables. These relationships could be positive or negative and the relationship varies in strength. OLS was chosen to minimize the error sums of squares, as that is the objective of this particular type of regression. Various assumptions are needed to effectively use the OLS estimator. The first assumption is the model is linear in parameters. This means that the dependent variable is a linear function of independent variables and the error term. Secondly, the number of observations must be larger than the number of parameters in the model. The third assumption is the sample of observations must be random as to not have any biases in the data. Assumption four is conditional mean should be zero meaning there must be no relationship between the X's and the error term. Next is the fifth assumption of homoscedasticity meaning all of the error terms in the regression have the same variance. Assumption six is no auto correlation between the error terms and the seventh and final assumption is no multicollinearity. Multicollinearity is correlation among explanatory variables. Given these assumptions, the model for this researched is expressed as:

(1) Illinois Soybean Yield = β_0 + β_1(average row width) + β_2 (percent biotech acres) + β_3 (percent double cropped acres) + β_4(June precip.) + β_5(July precip.) + β_6(August precip.) + β_7 (September precip) + β_8(June temp.) + β_9(July temp. + β_{10}(August temp.) + β_{11}(September temp.) + β_{12}(trend) + ε_i

In equation (1), Illinois soybean yield is set as a function of the ten independent variables. The expected sign of the coefficient for average row width will be negative. Given figure 6, decreasing row width is in conjunction with increasing yields. The coefficient for percent biotech acres will be positive. An increase in acres planted with GMO soybeans should result in an increase in soybean yield. This is also illustrated in figure 6 where the lines are along the same trend. The expected sign for percent double crop acres could go either way. With figure 6 showing a varying line for the variable, it is hard to decide what direction that will go. The expected sign for June precipitation, July precipitation, August precipitation and September precipitation is positive. Using a logical approach, it is believed that rain would increase the soybean yield, at least up to a certain point. However, it is hypothesized that June temperature, July temperature, August temperature and September temperature will all have a negative effect considering hot weather impedes soybean plant growth. Once again, this is only true to a certain extent. Weather too cold would also impede plant growth. Lastly is the variable representing the trend. With this variable representing all other factors influencing soybean growth, it is hypothesized that this will have a positive coefficient. There are many other factors including seed variety, fertilizer consumption, weed pressure, machinery used, and so much more. These crucial influences are also likely to positively impact soybean yield, however they are difficult to find data to represent them in this research.

In running this regression model, we will use many approaches to interpreting the effect of each independent variable on the dependent variable of soybean yield in Illinois. We will interpret the estimates to estimate the effect of each variable on yield; next we will conduct a t-test and then use T-statistic to determine significance or insignificance of a variable. Following that, R-squared can be used to determine how much of the variation in soybean yields is explained by all of the variables as a whole. Likewise, an F-test will explain variation in soybean yield as explained by my ten variables. Ultimately elasticities will be tested to determine the influence of each variable on soybean yield. Hypothesis tests for this study can be expressed as:

Table 2: Hypothesis Tests

Null Hypothesis	Alternative Hypothesis
H_o: $\beta_{AverageRowWidth} = 0$	H_a: $\beta_{AverageRowWidth} \neq 0$
H_o: $\beta_{PercentAcresBiotech} = 0$	H_a: $\beta_{PercentAcresBiotech} \neq 0$
H_o: $\beta_{PercentDoubleCropAcres} = 0$	H_a: $\beta_{PercentDoubleCropAcres} \neq 0$
H_o: $\beta_{JunePrecipitation} = 0$	H_a: $\beta_{JunePrecipitaiton} \neq 0$
H_o: $\beta_{JulyPrecipitation} = 0$	H_a: $\beta_{JulyPrecipitation} \neq 0$
H_o: $\beta_{AugustPrecipitation} = 0$	H_a: $\beta_{AugustPrecipitation} \neq 0$
H_o: $\beta_{SeptemberPrecipitation} = 0$	H_a: $\beta_{SeptemberPrecipitation} \neq 0$
H_o: $\beta_{JuneTemperature} = 0$	H_a: $\beta_{JuneTemperature} \neq 0$
H_o: $\beta_{JulyTemperature} = 0$	H_a: $\beta_{JulyTemperature} \neq 0$
H_o: $\beta_{AugustTemperature} = 0$	H_a: $\beta_{AugustTemperature} \neq 0$
H_o: $\beta_{SeptemberTemperature} = 0$	H_a: $\beta_{SeptemberTemperature} \neq 0$
H_o: $\beta_{Trend} = 0$	H_a: $\beta_{Trend} \neq 0$

CHAPTER 5

RESULTS

The regression results are shown in table 1 of the appendix. These coefficients are interpreted relative to the average Illinois soybean yield in bushels per acre for each year. First and foremost, the estimates are interpreted using the estimated coefficient from table 1 in the appendix. These are defined as $\beta_{AverageRowWidth}$ = for every inch increase in average tow width, yield decreases by .005 bushels per acre. $\beta_{PercentAcresBiotech}$= for every 1% increase in percentage of acres biotech, yield decreases by .01 bushels per acre. $\beta_{PercentDoubleCropAcres}$= for every 1% increase in percentage of acres double cropped, yield decreases by 0.37 bushels per acre. $\beta_{JunePrecipitation}$= for every inch increase in June precipitation, yield increases by 0.44 bushels per acre. $\beta_{JulyPrecipitation}$= for every inch increase in July precipitation, yield increases by 0.88 bushels per acre. $\beta_{AugustPrecipitation}$= for every inch increase in August precipitation, yield increases by 1.78 bushels per acre. $\beta_{SeptemberPrecipitation}$= for every inch increase in September precipitation, yield decreases by 0.14 bushels per acre. $\beta_{JuneTemperature}$= for every degree increase in average June temperature, yield increases by 0.37 bushels per acre. $\beta_{JulyTemperature}$= for every degree increase in average July temperature, yield decreases by 0.12 bushels per acre. $\beta_{AugustTemperature}$= for every degree increase in average August temperature, yield decreases by 0.67 bushels per acre. $\beta_{SeptemberTemperature}$= for every degree increase in average September temperature, yield increases by 0.09 bushels per acre. And finally, β_{Trend}= for every unit increase in other factors (trend), yield increases 0.52 bushels per acre.

The next step is to define the significance or insignificance of each variable using a t-test. With degrees of freedom being 30 and level of significance being 0.05, the t critical value was established at -2.04 and 2.04. With that being said, every t statistic in the regression results that falls between the critical values we fail to reject. Likewise, every t statistic falling outside -2.04 and 2.04 is rejected. Four of the variables in the model are rejected: July precipitation, August precipitation, August temperature and trend. These variables are the ones showing a significant effect on yearly average yields. July precipitation, August precipitation, and the trend both show a positive significant effect while August temperature shows a negative significant effect. This effect can be shown in figure 6 of the appendix. An obvious correlation is apparent with high august temperatures causing a simultaneous decrease in yield. This can be seen in figure 5 of the appendix.

The next type of explanation that is crucial to this research is the R-squared interpretation. The R^2 shown in table 1 is depicted as 0.8503. This is described as 85.03% of variation in soybean yields can be explained by row width, percentage of acres biotech, percentage of acres double cropped, June precipitation, July precipitation, August precipitation, September Precipitation, June temperature, July temperature, August temperature, September Temperature, and the technology trend. 85.03% is a very good result for this test being high, but not too high. Very high R^2 coupled with low T-values results in multicollinearity, which this data does not show. Multicollinearity is simply independent variables correlating with each other.

In relation to the R^2, an F-test was conducted to test the R-squared. This is used to determine if R^2 is equal to zero. With degrees of freedom identifying as (12,30), level of significance chosen is .05, a critical value of 2.14, and F being calculated as 26.24, the hypothesis of R^2 being equal to zero is rejected. This agrees with the R-squared test that 89.12% of variation in soybean yield can be explained by the variables depicted in this paper.

CHAPTER 6

DISCUSSION

This study presents an analysis of variables impacting Illinois soybean yields for the last 42 years. These 42 years were chosen due to the availability of data and years with the chosen variables present. The data set includes data from 1974 to 2016 including the dependent variable, average soybean yield, along with twelve independent variables. They were analyzed using a multiple linear regression, which recognized four of the independents as significant. These include July precipitation, August precipitation, August temperature and the trend variable. It is surprising that the variables flagged as significant are due to the uncontrollable forces of Mother Nature. With farmers and their respective employers working each year to improve yields whether it be with new ideas in row spacing, variety changes or many other facets, the majority of the yield comes down to the temperature and precipitation the farm gets in a given year. However, with my trend variable the intention was to represent the increasing changes and developments of knowledge and technology on the farm. So some of these aspects could also have impact on soybean yield as well. With double crop, biotech, and row width all being insignificant, it is apparent that there are other reasons for these advancements including easier weed control and change of equipment. The implementation of herbicide tolerant crops was introduced to simplify crop protection and reduce weed pressure, not to improve yields.

Because this study only takes Illinois' data into account, it would be difficult and mostly useless to consider these data in other states. It also has some other variables that could be added in the future including fertilizer consumption and May weather variables. Despite these limitations, I believe the analysis gives a good idea as to what variables do not impact soybean yield the way the general population assumes. This data will allow farmers to become more educated on methods to improve yields.

BIBLIOGRAPHY

CMB.Contact@noaa.gov. "Climate at a Glance." National Climatic Data Center. Accessed April 18, 2017. https://www.ncdc.noaa.gov/cag/time-series/us.

Green, Jerry M. "Evolution of glyphosate-resistant crop technology." *Weed Science* 57, no. 1 (2009): 108-117.

Goldblum, David. "Sensitivity of corn and soybean yield in Illinois to air temperature and precipitation: the potential impact of future climate change." *Physical Geography* 30, no. 1 (2009): 27-42.

Gujarati, Damodar N., and Dawn C. Porter. Basic Econometrics. Boston: McGraw-Hill/Irwin, 2009.

Hymowitz, Theodore. "Introduction of the Soybean to Illinois." *Economic Botany* 41, no. 1 (1987): 28-32. http://www.jstor.org/stable/4254930.

"Illinois Soybeans Matter." Soybeans for Illinois. 2017. Accessed April 12, 2017. http://www.soybeansforillinois.org/index.html.

Sanders, Dwight R., Ira J. Altman, and Nick Ferraro. "Technological Advances in Soybean Yields: A Disaggregated Approach." *Agribusiness* 30, no. 2 (2013): 207-21. doi:10.1002/agr.21348.

"SOY AND ITS USES." WWF. 2017. Accessed April 12, 2017. http://wwf.panda.org/what_we_do/footprint/agriculture/soy/soyreport/soy_and_its_uses/.

Tannura, Michael A., Scott H. Irwin, and Darrel L. Good. "Weather, technology, and corn and soybean yields in the US corn belt." (2008).

USDA/NASS QuickStats Ad-hoc Query Tool. Accessed April 18, 2017. https://quickstats.nass.usda.gov/.

APPENDICES

APPENDIX

Table 1:

Model		Estimated Coefficient	Coefficients Standard Error	t - statistic	Hypothesis Test Outcomes
1	Constant (Average Yield)	47.7222	24.3847	1.9571	
	Average Row Width	0.0054	0.1494	0.0364	Fail to Reject
	Percent acres biotech	-0.0114	0.0297	-0.3836	Fail to Reject
	Percent acres double crop	-0.3683	0.3198	-1.1516	Fail to Reject
	June Precipitation	0.4364	0.2586	1.6874	Fail to Reject
	July Precipitation	0.8885	0.3642	2.4398	Reject
	August Precipitation	1.1778	0.3464	3.3996	Reject
	September Precipitation	-0.1401	0.3393	-0.4129	Fail to Reject
	June Temperature	0.3765	0.2772	1.3584	Fail to Reject
	July Temperature	-0.1203	0.2240	-0.5369	Fail to Reject
	August Temperature	-0.6676	0.1981	-3.3697	Reject
	September Temperature	0.0922	0.1973	0.4672	Fail to Reject
	Trend	0.5180	0.1360	3.8087	Reject
Model Summary					
	Model	**R**	**R^2**	**Adjusted R^2**	**Standard Error of the Estimate**
	1	.945	.893	.850	2.838

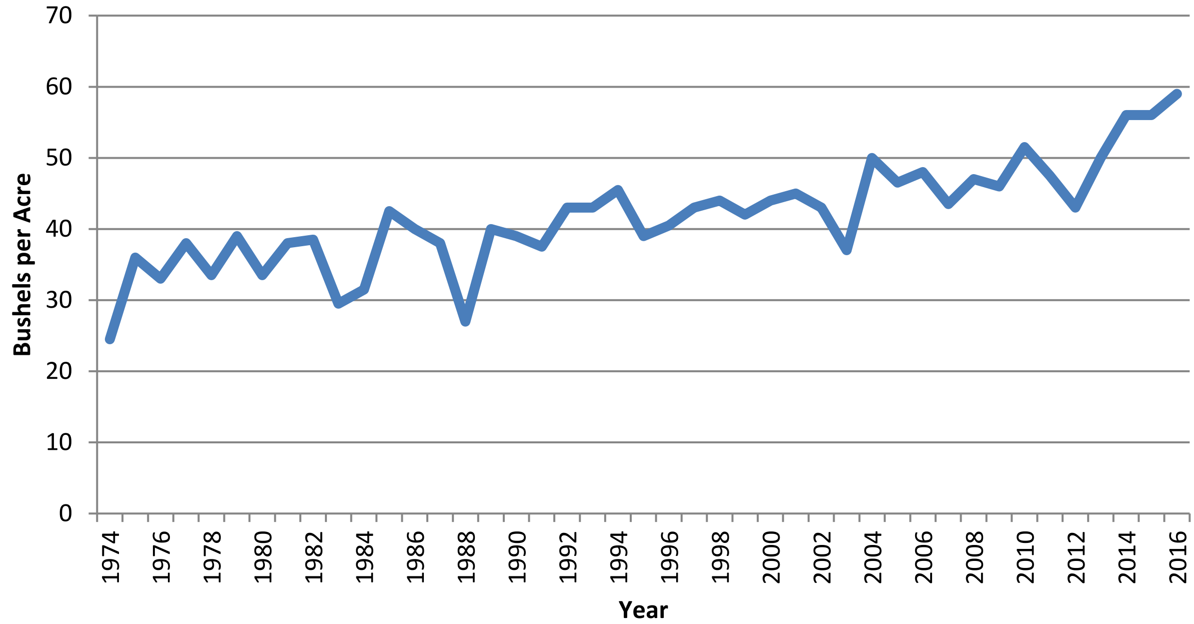

Figure 1: Illinois Soybean Average Yield 1974 – 2016

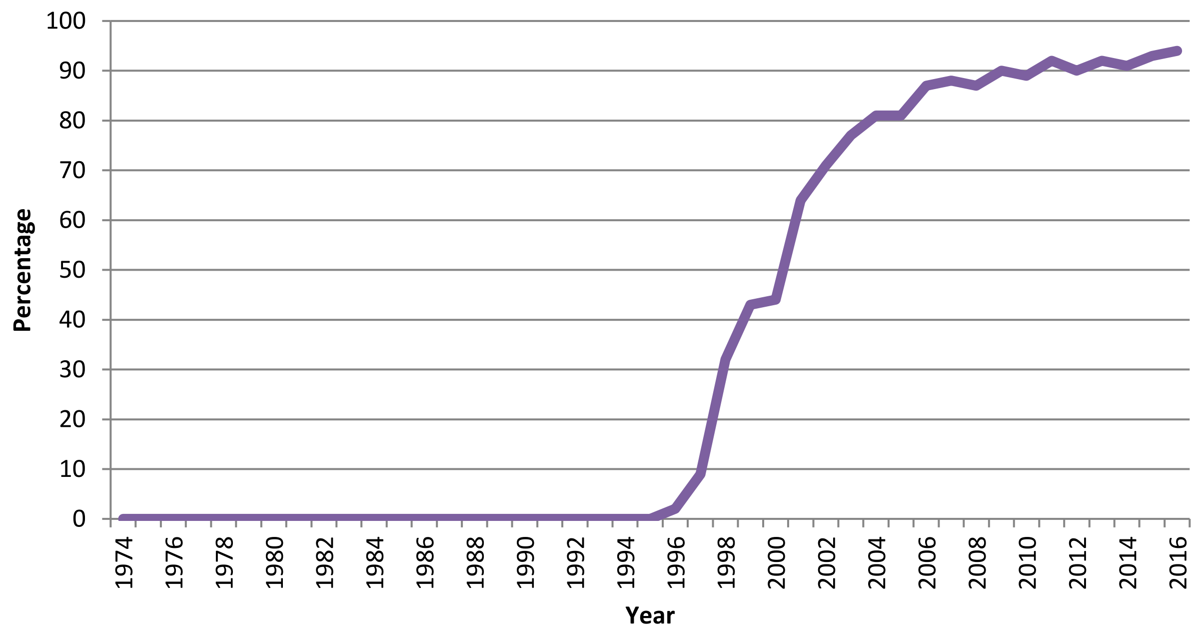

Figure 2: Illinois Soybean Biotech Percentages 1974 - 2016

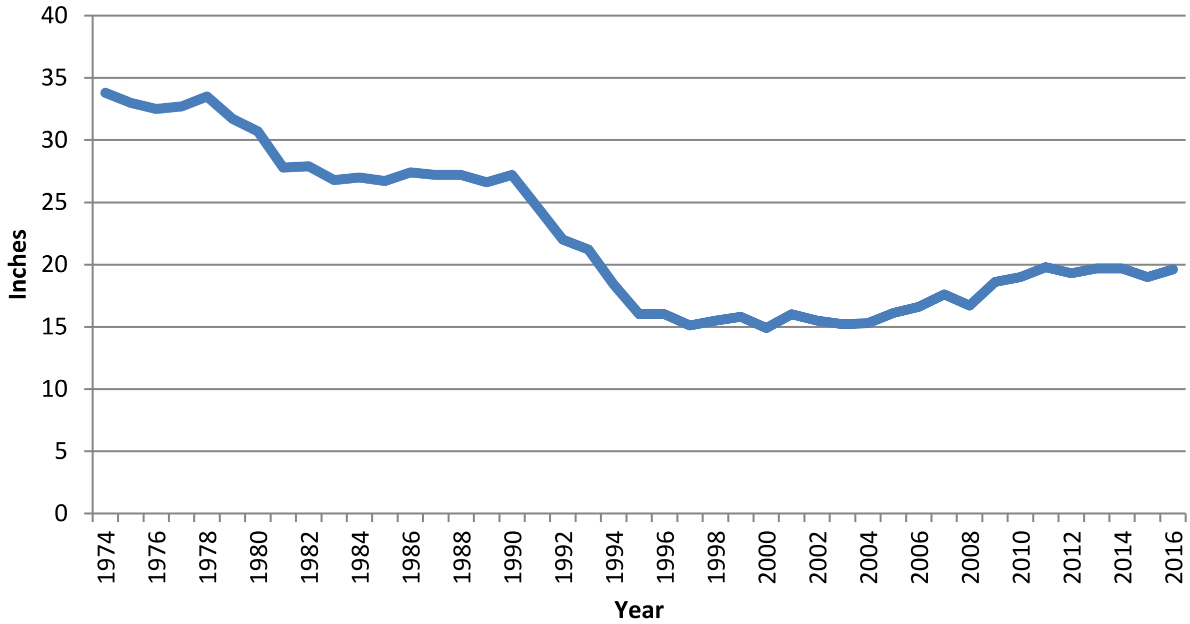

Figure 3: Illinois Soybean Average Row Width 1974 – 2016

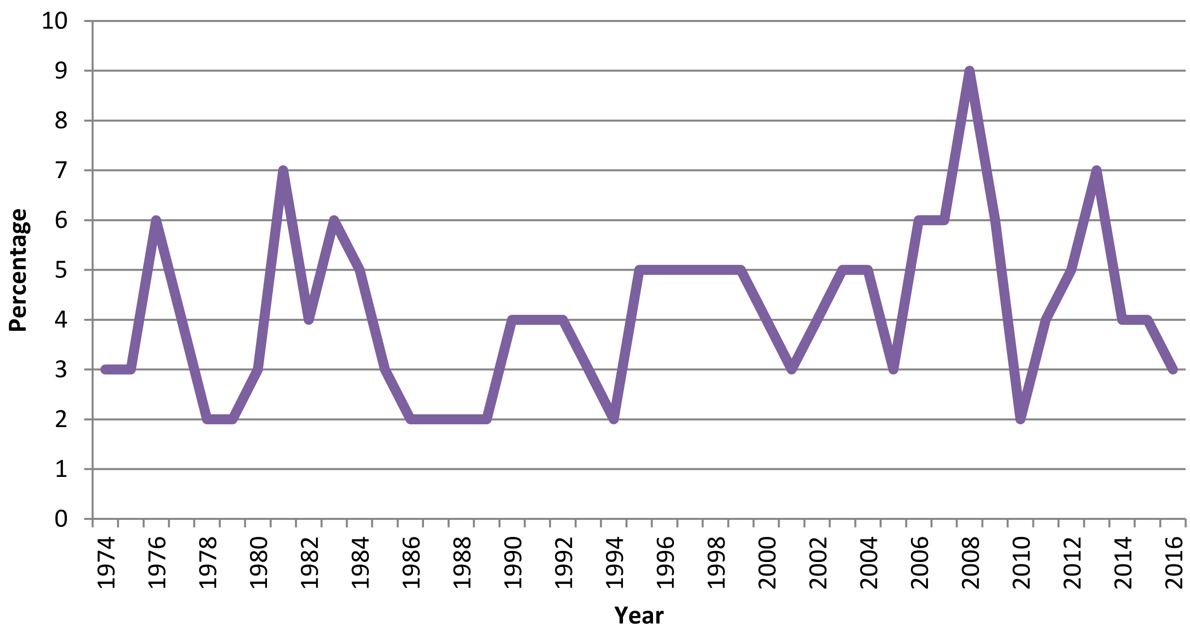

Figure 4: Illinois Soybean Double Crop Percentages 1974 - 2016

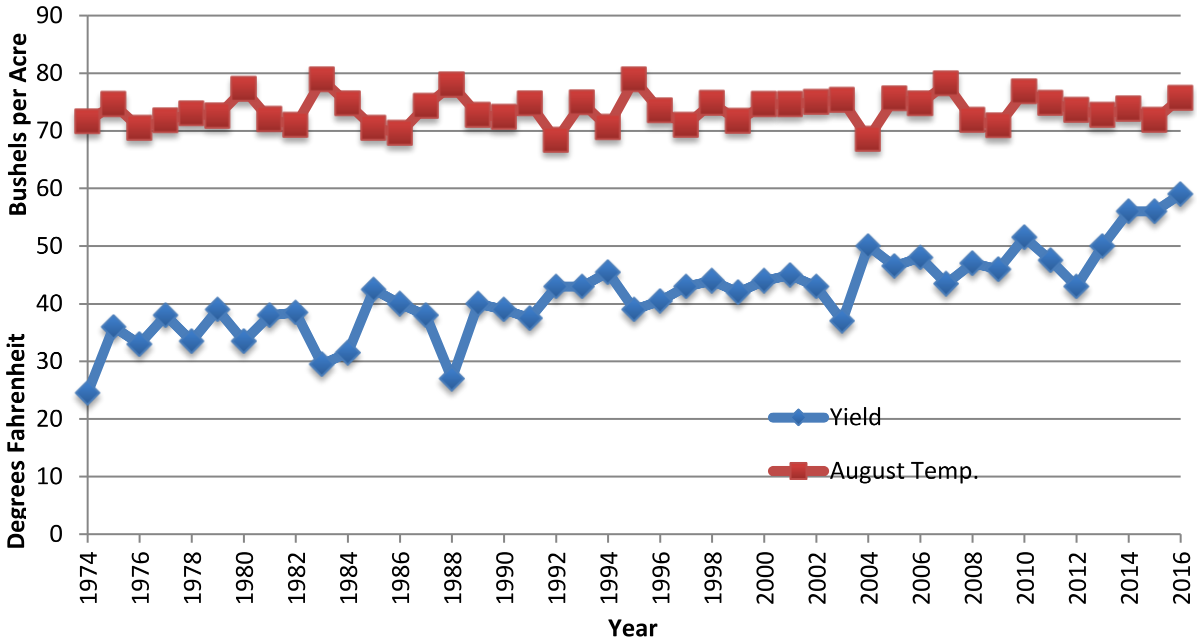

Figure 5: Illinois Soybean Yield / August Temperature Relationship 1974 - 2016

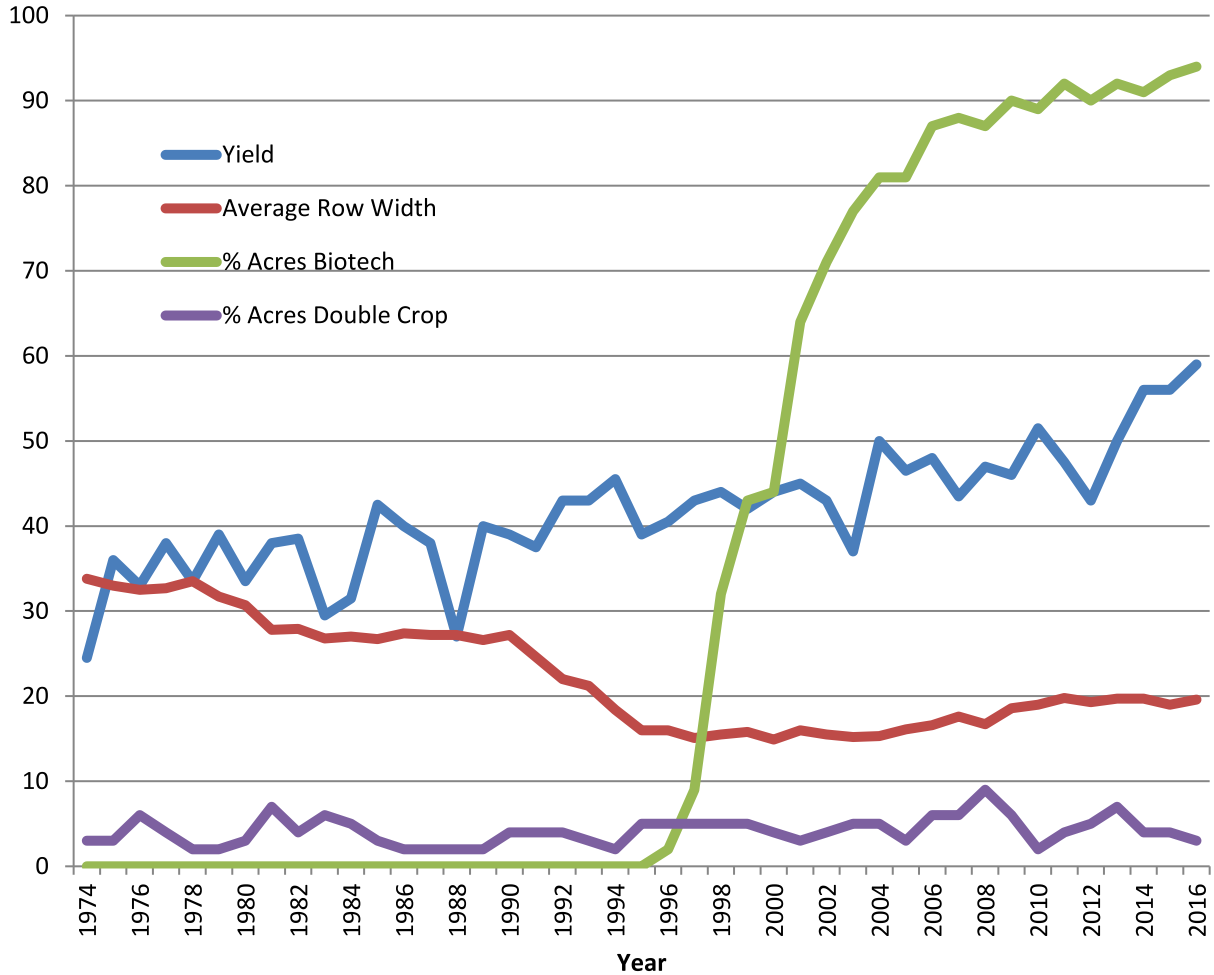

Figure 6: Illinois Soybean Yield / Technology Factors Correlation 1974-2016

VITA

Graduate School
Southern Illinois University

Sydney R. Bollinger

sydneybollinger@yahoo.edu

Southern Illinois University Carbondale
Bachelor of Science, May 2017

Research Paper Title: Factors Influencing Illinois Soybean Yield Increase

Major Professor: Dr. Dwight R. Sanders