

**THE DEVELOPMENT OF A CONCEPTUAL
BENCHMARKING TOOL REPRESENTING
BIG DATA AND AGRICULTURAL
TECHNOLOGY ADOPTION ON THE FARM**

by

JACOB LAFE MAURER

B.S., Kansas State University, 2009

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

2015

Approved by:

Major Professor
Gregory Ibendahl

ABSTRACT

One of the latest buzzes amongst agriculture is the storage and analysis of “Big Data.” There are a number of questions surrounding the quality, quantity, and capacity of big data to form real-world decisions based upon past information. Much like the teachings of history, the storybook that big data can reveal about a grower’s operation may hold the answers to the question of: “what is necessary to increase food production which will be required to feed an ever-growing world?” With the increase in interest in precision agriculture, sustainability practices, and the processing of the immense spatial dataset generated on the farm, the next challenge at hand will be in determining how to make technology not only streamlined, but also profitable.

Over the past few years, precision agriculture technology has become widely adopted as an agronomic decision making tool. Much like a scientific experiment, the greater the number of similar observations, the greater the degree of confidence can be placed upon a decision. As a means of increasing the number of observations that a farmer can use to base a decision upon, there is becoming increasing demand in being able to combine the data of similar farming operations in order to increase the size and scope of the dataset to generate better decisions benefitting many farms instead of just one.

The growing interest in forming community data pools for farm data demonstrates the need for a study for determining how farming practices can be properly benchmarked. The goal was be to evaluate how to use farm data to make economic decisions in a similar manner as one would make agronomic decisions using similar observations.

The objective was to design the proper protocol for benchmarking the farm's potential, and evaluating potential increases in technical efficiency by adopting precision agriculture technology. To accomplish this, a data envelopment analysis was conducted using scale efficiency as a means of determining the frontier of efficient farms.

The resounding goal for this study in the future will be to use the model as a means of implementing the secondary process of pooling precision agriculture data to analyze efficiencies gained by the adoption of technology. By demonstrating the value of generating peer groups to increase observations and refine farming practices, farmers can find increased profitability and efficiency by using resources that may already be held within the operation.

TABLE OF CONTENTS

List of Figures	vi
List of Tables	vii
Acknowledgments	viii
Chapter I: Introduction	1
Chapter II: Literature Review	3
2.1 Introduction	3
2.2 Precision Agriculture.....	3
2.3 Telematics and Telecommunications Infrastructure	7
2.4 Community-Pooled Data & Embodied Knowledge.....	8
2.5 Farrell’s Measurement of Productive Efficiency – Data Envelopment Analysis	10
2.6 Adoption of Data Envelopment Analysis in Agriculture	13
2.7 Re-Introduction of Data Envelopment as a Benchmarking Tool.....	14
Chapter III: Theory	17
3.1 Introduction	17
3.2 The Production Function.....	17
3.3 Price Efficiency versus Technical Efficiency.....	18
3.4 Scale Efficiency.....	19
3.5 Non-Parametric Data Envelopment Analysis (DEA)	19
Chapter IV: Methodology	21
4.1 Introduction	21
4.2 Outline of Project	21
4.3 Development of Dataset in R.....	24
4.4 Analytical Approach	30
4.5 Benchmarking in R.....	32
Chapter V: Data and Results	34
5.1 Data Introduction.....	34
5.2 Descriptive Statistics of Simulated Values.....	34

5.3 Scale Efficiency Analysis of Simulated Dataset	37
5.4 Scale Efficiency DEA for Commodity Sub-groups	39
5.5 Scale Efficiency DEA for Precision Agriculture Technology Sub-groups	41
Chapter VI: Conclusion.....	45
6.1 Future Work.....	46
References and Works Cited.....	48
Appendix A: R Script for Dataset Simulation	52
A.1 Introduction.....	52
A.2 Outline of Headers Found in Script	52
A.3 Dataset Simulation Script.....	52
Appendix B: R Script for Data Envelopment Analysis	57
B.1 Introduction.....	57
B.2 Dataset Analysis Script.....	57

LIST OF FIGURES

Figure 2.1: Farrell’s Isoquant Diagram.....	11
Figure 2.2: Farrell’s Scatter Plot	13
Figure 4.1: Distribution of Clay County Kansas Farms.....	22
Figure 5.1: DEA Plot – Scale Efficiency of All Farms.....	38
Figure 5.2: DEA Plot – Corn Producers	40
Figure 5.3: DEA Plot – Soybean Producers	40
Figure 5.4: DEA Plot – Grain Sorghum Producers.....	41
Figure 5.5: DEA Plot – Wheat Producers.....	41
Figure 5.6: DEA Plot – Adoption of Yield Monitors.....	42
Figure 5.7: DEA Plot – Adoption of Variable Rate Technology.....	43
Figure 5.8: DEA Plot – Adoption of Auto Guidance.....	43
Figure 5.9: DEA Plot – Adoption of Big Data Analytics.....	44
Figure 5.10: DEA Plot – Adoption of Telematics	44
Figure 6.1: Field Technical Efficiency	47

LIST OF TABLES

Table 4.1: Distribution of Acres as a Percentage of Total Simulated Acres by Crop.	26
Table 4.2: Distribution of Acres as a Percentage of Total Acres by Farm Size.....	26
Table 4.3: Distribution of Acres by Crop (Simulated Values).....	26
Table 4.4: North Central Kansas Average Yield 2005-2014	28
Table 4.5: Input Costs by Commodity	29
Table 4.6: Clay Center, KS 2015 New Crop Commodity Prices (03/13/2015).....	30
Table 4.7: Influence of PA Technology Adoption on Input Cost and Yield.....	30
Table 5.1: Descriptive Statistics (Simulated Values).....	35
Table 5.2: OLS Regression of Simulated Dataset.....	36
Table 5.3: Summary Statistics for Scale Efficiency Scores	38
Table 5.4: Censored Tobit Regression Model	39

ACKNOWLEDGMENTS

This project represents the culmination of many years of study. Through these years, both in the classroom and in industry, I have come to learn many things about myself and what is important in life. I have seen achievements which I thought were the greatest things in the world brought back down to reality in the blink of an eye, making me truly realize that it is the people with whom you surround yourself that make something special. This is my section to honor those people who have made this project special.

First off, I must thank the love of my life, Kendra. Her incredible resiliency has inspired me more times than I can count. She encouraged me to follow my dreams of returning to school, and even when I was rejected by several people and programs, she continued to push me. Her commitment to me empowered me to not just be another student, but to make this school experience something truly special. I mean, if you are going to do something you've always wanted to do, and you've had to fight the entire way, you are not going to do it half-way.

The same gratitude goes to my newfound mentor, Dr. Terry Griffin. It has been my distinct honor to have worked on this project with you and to continue to conduct research with you. What you have done for my career by simply coaching me and believing in me goes beyond the thank you I could ever give. Our late nights sending this document and the scripts back and forth made this process more than a simple project. It has become a part of my life and I am proud of what we have generated so far. The best is yet to come.

This project represents another generation of my family to be involved in production agriculture, and I am so incredibly proud to be able to continue on my family's

deep heritage in this industry. I feel that the inspiration for everything that I do in agriculture stems from my summers spent on the farm with my grandma and grandpa. They taught me the true meaning of husbandry and care for the land, and what it means to be a family man. For all of the times we sat and drank coffee in the mornings, ate ice cream bars on the swing out front, or spent hours on end in the tractor cab sitting on Grandpa's knee. Thank you. You both have made me who I am today.

To the members of my committee who have seen this project go from discussion to defense, I am proud to be presenting this for your review today. To the faculty of Kansas State University who have been a part of my education over these past three years, and to Gloria, Mary, and Deborah. With all of my heart, I thank you. Every individual has helped to mold me into the professional I am today and have more than helped to satisfy my hunger for learning. I am proud to forever represent you all as an alumni of this program.

CHAPTER I: INTRODUCTION

This thesis is designed to help with the demonstration of the technical efficiency associated with implementing precision technology and how that might help to answer the question of whether the technology brings profitability to the farm. The model to demonstrate this will be based upon the data envelopment analysis of an expansion of a dataset of limited observations. The purpose of using this type of dataset is to protect the private information of the growers of the survey used in this public forum. Upon the review of this work, the simulated dataset generated within this study will be replaced with real farm data.

The agricultural media has reported on ‘big data’ as the current buzzword over the last four years, however, very little success stories exist with respect to community data analytics. The overall goal of this thesis was to develop a commercially viable community analysis tool for benchmarking farmer-clients. This objective was specifically carried out by adapting existing data envelopment analysis (DEA) methods for use in real-time big data environment. The majority of farm-level ‘benchmarking’ is suspected to be informal ‘coffee shop talk’; and a more rigorous quantitative methodology has the opportunity to satiate the natural desire of farmers to rank and compare themselves with respect to yield output, input use, and more sensitive profitability.

Data envelopment analysis methods were applied to a representative dataset of farmers who have shared data with a local cooperative. These DEA results provided comparative analytics regarding farmers’ utilization of precision agriculture technology. The expected outcome of this study will be to determine the variables necessary to ensure a proper analysis can be conducted with the real data and to gather feedback from others in a

public forum to determine whether there are other means of calculating true technical efficiency from the adoption of precision agriculture technology on the farm.

This study will consist of four primary pieces. First will be a literature review to determine the proper method on which the analysis will be conducted. Secondly, a script will be written to expand the dataset beyond the limited observations using both 2012 USDA Ag Survey data and relevant works. Next, the dataset will be subjected to multiple forms of analysis to determine the proper format to calculate and analyze technical efficiency. Finally, the project will be submitted to the committee for discussion on improving the model and ensuring the analysis will be conducted properly prior to constructing a survey for data collection to build the database in industry.

CHAPTER II: LITERATURE REVIEW

2.1 Introduction

There is much debate among the experts within agriculture as to the necessary requirements of the modern farm operation that will allow for the growth needed to feed the ever-increasing population of the world. The techniques that come to mind deal much in improving conservation practices, increasing efficiency in equipment, and maximizing the potential yield of growing crops. While these seem to be a fairly logical set of requirements, there has not been nearly enough work done on the side of determining exactly how all of these methodologies will fit together to allow for proven sustainable growth on the farm. The purpose of this literature review is to outline some of the work that has been done in this part of the industry, with a strong focus on technical efficiency and its definition as well as discussing some of the current methodologies that are being discussed as potential sources of research for further work, of which this project is a part.

2.2 Precision Agriculture

Precision Agriculture, simply put, is information technology applied to agriculture (Whitacre, Mark and Griffin 2014). Within this basic outline come the subjects of big data, telematics, and embodied knowledge. Each of these methodologies encompasses the framework powering the technologies that serve modern day farmers and assist them in a variety of ways in making on-farm decisions.

Simply put, precision agriculture and big data can be seen by most as interchangeable. The differences do not have that much of an effect on the surface to the average end-user, but the fact is, the powers within big data are just that – big. What is meant by big data is more than a spreadsheet containing thousands of entries. It is more than bringing together a variety of spatial data and analytics to assist in making agronomic

decisions. Big data describes the process of growth and enhanced decision making analytics, being both spatial and intangible, that empower the end-user to fully incorporate all aspects of the operation to make customized, site-specific decisions. The power unleashed by being able to utilize a vast array of unbiased, original data points can assist a grower in making true-to-life decisions that are based upon any number of variables of his own choosing, further incorporating the grower's own personality into the decision making process.

Precision Agriculture (PA), on the other hand, is an explanation, or description of the tools and processes that further enhance common agricultural practices by adding an element of specificity (geo-referenced points). Many associate precision with GPS, or the Global Positioning System but really what they are referring to is a device using GNSS, which is a Global Navigation Satellite System (Whelan and Taylor 2013). Other devices and methods that are found within the framework of PA are Unmanned Aerial Vehicles (UAVs), electrical conductivity (EC) or Veris[®] data, yield monitors, satellite imagery, and variable rate technology (VRT). The majority of these devices have a sole purpose of providing the grower with feedback regarding field conditions, soil structure, and production linked by a geographical reference point to allow the overlay of multiple layers of data for analysis.

A newer technology to hit mainstream agriculture is that of telematics. This technology involves the wireless transfer of data between multiple devices or multiple machines. This automated transfer of data is accomplished using cellular technology in most cases as a data carrier to connect to cloud-based data storage points. The adoptions of this technology generally occur through the purchase of new equipment such as tractors,

combines, and spray rigs, but can also be added to old equipment as well as a stand-alone device tied into the machine's on-board computer system. The systems transmit not only as-applied and yield data to the cloud, but also enables the user to view real-time telemetry regarding the machine's location, speed, and direction as well as engine fluid levels and temperatures in some systems.

Mark, Whitacre, and Griffin (2015) outlined a three-stage adoption process that a typical farmer might follow in implementing precision technology in their operation in a paper presented at the Southern Agricultural Economics Association's Annual Meeting. Stage one of the process is the purchase of a yield monitor and beginning the practice of collecting yield data. Even though in the paper they state that this process is ubiquitous due to the relatively inexpensive cost per observation and the fact that most new combines have yield monitors installed as a standard equipment option (a very true observation in operations with economies of scale and who are upgrading equipment), the initial overhead cost may be the hardest step to overcome for a small to middle-sized farming operation where the purchase of a new combine may not be feasible. It may become cost-effective to purchase the yield monitor out-right, but purchases such as this can be difficult for a grower to realize a net-return due to the fact that this piece of equipment is non-essential to the growing of crops in a non-precision farming operation. To combat this, some Kansas farmers instead start on the side of managing inputs rather than collecting data regarding production. In economic theory, which will be better outlined in later chapters, this may not be the worst alternative. The limiting factor, however, is the inability to measure true production spatially, and thus being able to determine whether site-specific product placement truly produced a positive return on their investment. This will become a key

limitation in being able to measure true growth as a result of PA, and because it takes at least three years of quality, normalized yield data to begin to make properly-quantified fertilizer and planting prescriptions, time may end up being the most daunting obstacle that growers will have to overcome in order to further progress in PA.

Stage two is noted as increasing the usage of soil mapping technologies. This includes, but is by no means limited to, EC mapping, grid sampling, and the usage of spectral reflectance sensors. Again, steps one and two may be reversed in some operations due to access to capital, poor commodity market conditions, and high input costs. Also, some agriculture input retailers may have programs to incentivize growers to implement different technologies or agronomic management programs that may allow for an increase in production which could possibly supplement the investment in equipment upgrades.

Finally, the third step in the process is the adoption of variable rate technologies (Mark, Whitacre and Griffin 2015). If a grower is outsourcing their PA to a retailer, this step may have been implemented sooner rather than later in the process. Because many custom applicators now have the necessary controllers as part of their standard equipment and they have greater economies of scale in the application of fertilizer spatially, this technology could be incorporated in even a small farming operation's agronomic practices many years before yield data will be accessible.

A farming operator's ability to grow within PA will have a direct impact on the measure of the farm's technical efficiency. A greater level of PA found on a given field may not only show greater returns at harvest time, but also greater transparency, which will become valuable as sustainability and traceability practices evolve and record keeping becomes a more integral part of the food supply. This is all summed up quite well by

Schimmelpfenning and Ebel (2011) in a USDA study quantifying the adoption of PA in US Agriculture. The conclusion of the research shows that the future viability of PA will depend upon three items: whether technologies become less expensive and/or easier to install and maintain, whether conservation tillage becomes more widespread, and whether the prices of fuel, fertilizer, and custom application are relative.

2.3 Telematics and Telecommunications Infrastructure

Telematics represents the next generation of technology to be implemented on the farm. Essentially, the advantage of such an adoption on the farm would be to eliminate the need to transfer data between devices because the communication will be seamless and require nothing more than a high-grade broadband source. As of right now, the data being transferred is not requiring a great amount of bandwidth because the number of devices and operations that currently implement this technology is not very wide spread. However, as the number of farmers adopting telematics for data transfer grows, the capabilities required of the telecommunications infrastructure will need to continue to improve in order to handle the increased upload demand brought on by agriculture.

There are four sources in which farmers obtain high-speed internet: via wired broadband (cable, digital, or fiber-optic line), dial-up (over telephone lines), satellite, and cellular service. The current debate deals specifically with the modern-day and future farm operator's telecommunications needs moving forward with regard to infrastructure and capacity to handle big data.

It may come as no surprise that urban areas have the strongest and most capable wireless network. Actually, 100% of the urban populous has access to at least one network provider. While 98% of the rural population is in the same position, the question no longer deals nearly as much with access as it does with signal strength and upload speed. As

Whitacre, Mark and Griffin outlined via multiple sources of their own: “only 90% of households have access to mobile wireless speeds of 3 Mbps or greater and only 78% have access to 6 Mbps or greater.” What does this mean for agriculture? If not even all urban areas have access to high-speed internet, and the demand by rural residents continues to grow as technology expands the usage of wireless devices to farm implements, then the needs of the modern farm could make the argument for upgrades in telecom infrastructure much greater than anticipated. The questions that need to be answered are: what is the economic efficiency threshold that comes between agricultural growth and broadband speed, and will the farm of the future have the necessary access to data transfer speeds that will allow them to evolve in the twenty-first century and to adapt the new technology being outfitted in new equipment (Mark, Whitacre and Griffin 2015)?

What is needed is to define technical efficiency as a means of how the farm is able to grow, or not grow, by adapting precision technology methods – namely telematics. Also, it will be necessary to determine the threshold which exists where growth in this field will outweigh net farm income. It will be critical to then demonstrate the current capability as well as the missing links that will keep from expansion and growth in the usage of telematics in big data.

2.4 Community-Pooled Data & Embodied Knowledge

In its current form, precision technology requires a substantial commitment of both capital and time in order for a farmer to become fully-integrated, which could be a reason that such technologies have not been adopted in lower-yielding environments. The capabilities that exist within the technology and the increased availability to automated systems will continue to have an impact in this facet of agriculture and will help in the growth of its use.

In time, precision technology will completely change the way farm data and records are stored. It will also change the means of comparing farmers amongst their peers. Where the status of a farm was compared to those within a neighboring geography and based upon yields and equipment, today, the existence of cloud-based storage, community data pools, and data cooperatives allow for the comparison of farms among potentially millions of others who may be in completely separate parts of the world. The ability for growers to share experiences and techniques with a group beyond their immediate geography could enable them to expand their knowledge base and to possibly find new management practices to enhance profitability.

This enhanced data pool could help growers help themselves, but could also provide direction for retailers and cooperators in targeting new methods to research, and catering programs toward solving problems existing among self-segmented peer groups. The more complete the dataset of a community, the more realistic models can become to provide insight and direction for everything from research and development of new products to risk management practices. The enhanced data could also become an additional tool to be utilized by those involved in other aspects of farming other than crop production. It could be used to generate realistic models for grain storage and merchandising groups, to enhance crop production forecasting, or even to assist in product allocation for agriculture retailers.

As more growers become interested in incorporating PA on the farm, the more that must be known about its use and how the farmer wishes to use the information. While the “big picture” involves heavy involvement of big data processing, the ability for the average end-user to put this to practice will allow them to adapt the added process into the current

day-to-day activities without requiring them to make any drastic changes. There becomes a need, then to designate how the process can be separated between an information-intensive process and embodied knowledge (Griffin, et al. 2004).

Embodied knowledge, as described by Whitacre, Mark, and Griffin in the *Choices* article, involves a purchase where information is passed to the end-user in the form of an input which will require no acquisition of additional skills in order to adopt the advanced technology. Such examples from the article include automated steering and swath controls in equipment and Round-Up Ready[®] technology in seed. When an end-user purchases this technology, no increase in skill set is required. Since the technology package is ‘self-contained’, the customer is able to adapt the enhanced capability without a need to purchase any additional equipment or change the way the product would otherwise be used within the operation (2015).

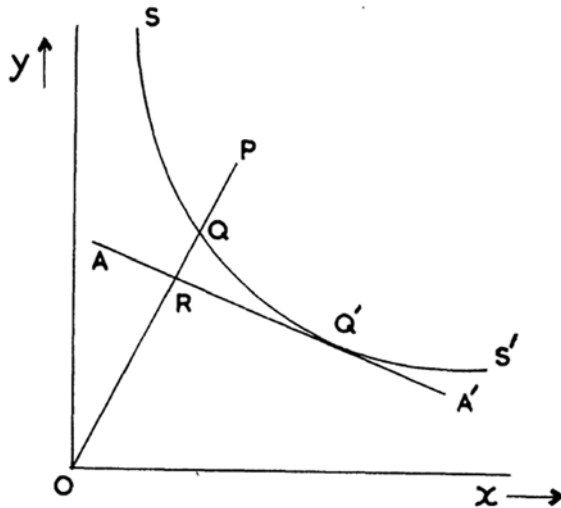
2.5 Farrell’s Measurement of Productive Efficiency – Data Envelopment Analysis

To properly discuss the classical process of Data Envelopment Analysis (DEA), a review of the work done by Farrell should be included for his insights on how technical efficiency measurement in this form came to exist. In one of the first essays dealing purely on determining a measurement of productive efficiency, taking into account more than just labor in its theory, his work outlined a number of concepts that are relevant to nearly every DEA application.

In Farrell’s example, he begins with an exercise outlining a simple case of a technically efficient firm which employs two factors of production that will result in a single end-product, produced under conditions of constant returns to scale. The efficient production function is known, which in essence means that the value of the perfect combination of inputs is known where output will be maximized.

Farrell's diagram (Figure 2.1) graphically illustrates where the point P represents the inputs of the two factors, per unit of output, that are used in production. The isoquant SS' is representative of the variety of combinations of the factors representing a perfectly efficient firm. Point Q represents a second efficient firm using the two factors in the same quantities as P . This means that the second firm, Q , is able to produce the same amount of output as P , but at a fractional rate OQ/OP times that of P . Where the original theory was to simply be satisfied that OQ/OP was indeed the technical efficiency of firm P , Farrell redefined technical efficiency. He also took into account factors related to optimum production in a given marketplace, where price becoming a part of the equation is just as important as the output.

Figure 2.1: Farrell's Isoquant Diagram



Source: (Farrell 1957, 254)

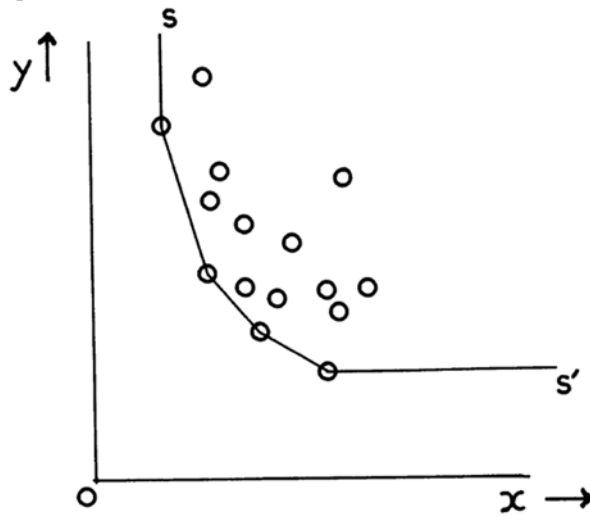
If AA' (Figure 2.1) has a slope equal to the ratio of the prices of the two factors, the optimal method of production would, in fact, not be Q , but rather, Q' . The rationale being that even though both points represent a technical efficiency of 100 percent, a firm producing at point Q' will be producing at a fraction of the cost of Q equal to OR/OQ . The

end result in his example was that if the firm were perfectly efficient, by taking into account both technical, and also, price efficiency, the costs would be reduced by a fraction of *OR/OP*.

Farrell was quick to point out that there were a number of weak points in this simple assumption, noting that one would need to take into account the difference between theoretical and empirical decision-making and the fact that the theoretical efficient function overlooks a number of very complex parts to the process, essentially what makes firms different from one another. To this, the design of the model will depend upon the complexity and degree of realism. He also added that there are two very important assumptions that need to be made in terms of the design of the efficient production function: the isoquant is convex to the origin and in no way has a positive slope.

The efficient production function is a result of the two assumptions noted above combined with the individual firms on the scatter plot (Figure 2.2). The assumption of convexity comes into account because the resulting theory is, if any two points on the plot, or firms, are able to be represented in practice, then too, will any point which represents a weighted average of the two. The second assumption, being that of slope, exists because if not, an increase in the applications of both factors would result in reduced output. In the end, the curve *SS'*, becomes the estimate of the efficient isoquant.

Figure 2.2: Farrell's Scatter Plot



Source: (Farrell 1957, 256)

The resulting curve is then defined in the following way. There are line-segments that join the outer-most points in addition to the points representing infinity along each axis. Each point on the scatter plot represents the solution to the equation where $P_i = (x_{i1}, x_{i2})$, and where λ_{ijk}, μ_{ijk} represents the solution to the equations $\lambda x_{i1} + \mu x_{i1} = x_{k1}$ (likewise, the same for firm 2 and so forth). The solution to $\lambda x_{i1} + \mu x_{i1}$ will always be equal to, or greater than 1 for all P_k in the dataset A . The equation for technical efficiency of firm P_k is the maximum of $1/(\lambda_{ijk} + \mu_{ijk})$ for all segments $P_i P_j$ of SS' (Farrell 1957).

2.6 Adoption of Data Envelopment Analysis in Agriculture

In the mid-1980s, there were a number of studies conducted that began to introduce non-parametric analysis as a means of exploring technical efficiency in production agriculture. By the mid-1990s, Allen Featherstone and a number of Kansas State University academics had developed a variety of studies putting to use the modernized version of DEA to model and explore on-farm decisions and their impact on efficiency. One of the earliest studies, by Moghnieh, Featherstone, and Goodwin, introduces the “augmentation hypothesis” as a means of modeling technical change. Simply put, the

hypothesis states that “technical progress increases the effectiveness of inputs in the production of output” (Moghnieh, Featherstone and Goodwin 1991). The difference in this study versus others prior was that it recognized the fact that there was two different means in which a firm would determine the way it conducts business activities. A firm would either seek to maximize profitability or to minimize costs, the decision being made among a host of technological advantages and pricing options.

In a 1996 study, Featherstone and Rahman used a non-parametric analysis as a means of analyzing the profit-maximizing behavior of agricultural cooperatives. Because of the nature of the cooperative as a non-profit organization, it is assumed that the profit-maximizing model for this group would be different from that of a privately-held or publicly-traded company. Their analysis involved comparing 20 cooperatives across 8 states in the Midwest which were affiliated with Farmland Industries and determining the number of violations to the profit-maximizing hypothesis each cooperative held. From there, they were able to determine how much profit was forfeited as a result. In the end, they were able to use the non-parametric analysis as a means to compare against companies that, while belonging to a common group in Farmland Industries, otherwise have their own means of generating revenue and managing expenses (Featherstone and Rahman 1996).

2.7 Re-Introduction of Data Envelopment as a Benchmarking Tool

In 2006, Fleming et al. re-introduced the medium of comparative analysis to agriculture in the form of benchmarking. They noted five distinct criticisms to using DEA as a method for comparative analysis:

- “It failed to incorporate sound economic principles in its application.”
- “There was limited scope for action once indices were calculated.”

- “The approach failed to establish causal relations between farming practices and performance.”
- “It was not consistent with a holistic approach to farm decision making.”
- “Risks and uncertainty in farm decision making were neglected.”

(Fleming, et al. 2006, 3)

In the original criticism, the reasoning behind the debate came because the analysis held no point of reference to anything other than the entities being compared. When the analysis is looked at as the initial point and research is conducted based upon the findings, the ability to act becomes feasible. Due to the nature of each entity in agriculture acting so differently, with a variety of practices and methodologies coming into play, it is often difficult for those in the community to find value in establishing a reference point amongst peers. This is where the inability of comparative analysis to establish a cause for the differences between entities enters the conversation. Methods such as grouping or segmentation were among the only means for such analysis to take place. Due to the differing nature of farming practices amongst even similar pairings, the ability to use this methodology was scarce.

Fleming et al. sought to use data envelopment as an effective tool in comparative analysis, but found that in order to use it they had to overcome several key issues. One issue included the perception that it is often difficult to use DEA as an effective method for economic study due to the difficulty to define a definite course for action. Their argument to that is if the benchmarking terms are designed using economic principles from the start, the scope for action is broadened, a holistic approach is included, risk is incorporated, and relationships could be more easily identify changes required. If those elements could be

incorporated, then a comparative analysis can indeed provide value. Including reference to the text of Pidnyck and Rubinfeld (2005, 595) define a firm as being technically efficient “if the output of one good cannot be increased without decreasing the output of another good.” Using this concept, Fleming et al. argue that the producer is in essence “adopting ‘best-practice’ production methods for a given production technology, and all points on the production contract curve represent technically efficient combinations of labor and capital.” They also add that the usage of DEA provides the ability for comparison of inefficient farms to identify peer groups, allowing for the design of a course of action for growers to improve (Fleming, et al. 2006).

As a means of countering the benchmark argument, Quintana-Ashwell and Featherstone introduced a new study in early 2015 using DEA as a means of demonstrating farm productivity growth. Their work investigated the growth measured by the Malmquist Productivity Index, which could be regarded as the existing benchmark for the Kansas Farm Management Association farms represented in the study. They decided to break down the index value to examine the primary sources of growth in productivity. Their research found technical change to be the driving factor, which differs from the research previously conducted, where relative prices were found to be at the core of the impact upon technical change and efficiency. The result, in their opinion, was that perhaps farm improvements and farm profitability resulted in an increase in input prices (Quintana-Ashwell and Featherstone 2015).

CHAPTER III: THEORY

3.1 Introduction

This project combined multiple economic principles to establish a benchmark for further analysis. It required the understanding of an individual grower's production function and how it relates to increasing output via changes in inputs. It also involved determining the best means of comparing firms by their ability to manage the allocation of scarce resources as outlined by Farrell via three distinct types of efficiency: price, technical, and economic. This study included the measurement of scale efficiency by conducting a non-parametric data envelopment analysis to establish a benchmark for the purpose of ranking and grouping farms based on their efficiency. These concepts combined to define a hypothetical model that was built with the intention of expanding to include empirical data with the end-result of designing a platform for the growth of community-pooled data and embodied knowledge.

3.2 The Production Function

This study essentially covers how the production function changes the technical efficiency of the farms of Clay County, Kansas. To begin, it is important to outline what is being meant by the production function and how it relates to this project. The production function represents the necessary combination of inputs which are required to produce one unit of output. Multiple combinations are positioned in an array across on a graph to make up what is known as the production possibilities frontier. This frontier represents all combinations of inputs and technological advantages that combine to create desired levels of output. Though not always the case, the production frontier generally has a curved shape, representing diminishing marginal product. What is meant by this is that the amount

of effort and/or inputs required to produce smaller quantities of output is less than what will be required at greater output levels.

In agriculture, as it is with many other industries, there is an optimum level of output that can be achieved that will have the greatest return, and at that point, any output level higher than that level will actually cost more per-unit than the output level which precedes it. Precision agriculture technology can assist in determining the optimum level of output which will maximize the efficiency of inputs. This study attempts to discover how the relationship of inputs combined with various levels of access to precision agriculture equipment affects the ability of many firms to maximize technical efficiency.

3.3 Price Efficiency versus Technical Efficiency

Many studies conducted come with a list of assumptions that must be made in order for a given theory or strategy to be tested across a variety of differing decision-makers. This study is no different. Because each farm operates on its own individual marketing plan, with its own profit schedule, and containing a number of different goals for output and efficiency operating on with its own production function, it is imperative to list in advance the variables being considered to have an impact on the study and which ones are deemed heterogeneous enough to other entities being compared that the variability need not be included within the study.

Many farmers, in their actions, will dispute this economic theory by claiming that, because they have the ability to determine the best price for their input purchases, and because they will also sell at any given point in time of their choosing that they have the ability to control the price in their efficiency frontier. The trouble with this is that each farmer is one of many who are in the business of growing a commodity. To this, each farmer is a price taker. An increase in production by this one farm will in no way affect

global prices in any facet, nor will a hold-out in purchasing result in a decrease in the cost of inputs. Farrell describes that it is nearly impossible to become truly price efficient and to know future price structure, and because of this (especially in the position of a price taker), the best way to increase efficiency is by the means of increasing technical efficiency through management practices (Farrell 1957).

3.4 Scale Efficiency

The data envelopment analysis uses two primary concepts as a means of determining the efficiency frontier. In the analysis conducted, scale efficiency is being calculated as the ratio of the data envelopment of the Net Farm Income as it relates to Acreage at variable returns to scale (VRS) and constant returns to scale (CRS). The primary purpose of calculating this number is to benchmark the farms within the study based upon this ratio.

3.5 Non-Parametric Data Envelopment Analysis (DEA)

There are two distinct approaches in the analysis of the production function across multiple entities or multiple methods. First, there is the parametric approach. Within this method, the data is manipulated into an optimized functional form and analyzed using a traditional, parametric technique to estimate any unknown parameters from an observed dataset. Firms are then evaluated side by side to make observations in their processes due to the fact that each firm is different and may operate on completely separate terms even if producing the same outcome. The other form is the non-parametric approach. As noted by Featherstone and Rahman, “the nonparametric approach examines the behavioral-optimization hypotheses without specifying any functional form for production technology” (1996, 267). By not specifying a functional form, an analysis is free to be conducted without the fear of a potential aggregation bias. These biases exist because no

two firms are the same, and factors differentiate them all due to technological advantages or differing input-sourcing or commodity-selling programs. The existence of these biases is among the limiting factors in the analysis of neoclassical optimizing behavior (Moghnieh, Featherstone and Goodwin 1991).

CHAPTER IV: METHODOLOGY

4.1 Introduction

The objective of this project was to demonstrate the level of technical efficiency brought to a grower's operation by implementing various levels of precision technology. This was evaluated by developing a hypothetical dataset based upon real observations expanded to fit the distribution of the 2012 Ag Census of Clay County, Kansas and comparing the results of the enterprises via data envelopment analysis. Using a non-parametric data envelopment analysis, the goal was to determine what pieces of equipment and technology bring the most value to the grower when it comes to being technically efficient based upon the principles defined by Farrell and Featherstone.

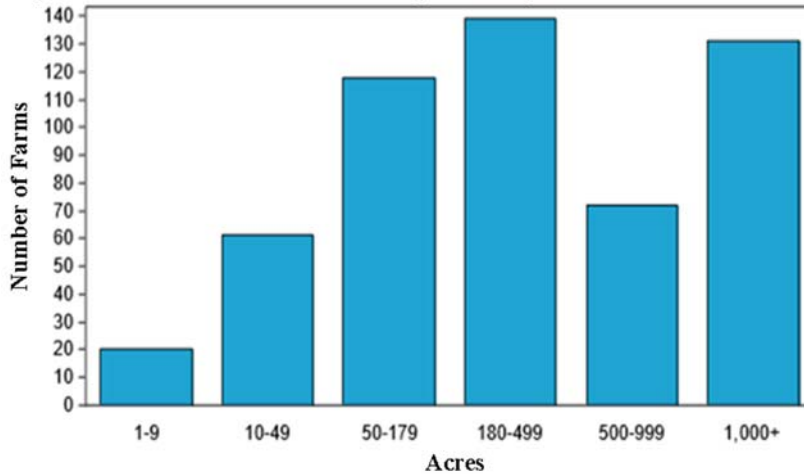
4.2 Outline of Project

The project consists of a script, written in R (R Core Team 2015) to simulate a county containing 541 farms to be representative of the distribution of farms in Clay County, Kansas in the 2012 Census of Agriculture (Figure 4.1). Using a truncated normal distribution, the farms of a similar distribution were generated with the objective of using the data to establish a number of simulated variables on which a data envelopment analysis can be conducted to measure technical efficiency. Upon establishing a distribution of farms, the acres were divided among a number of properties simulating a similar distribution to Clay County in 2012 (Figure 4.1).

The acres of each of the 541 farms were distributed into four commodity categories: corn, soybeans, sorghum, and wheat. The percentages of the total bushels delivered to Clay County elevators were 12.5% corn, 40% soybeans, 35% wheat, and 12.5% sorghum in the 2014 crop year. The acreage of each farm was simulated in the following manner, where each commodity's allocation was divided according to parameters placed upon farm

size. To begin, growers with acreage less than 1000 acres equally divided farm operations among soybeans, sorghum, and wheat. The allocation of acres for farms over 1000 acres was handled differently, where the distribution of commodity acres consisted of four separate computations.

Figure 4.1: Distribution of Clay County Kansas Farms



Source: (US Department of Agriculture 2012)

First, because corn was allocated only to this group of farms, there was a limiting factor in the form of an if/then statement which separated farms into the two separate groups and then applied a normal distribution of absolute values which represented the corn acreage to the group greater than 1000 acres. The end result was that 12.5% of the total acres in the simulation were allocated to corn, 100% of which originated from farms greater than 1000 acres in size.

Next, the total number of acres of each farm over 1000 acres was multiplied by 0.43, representing the remaining acres of soybeans to be represented by farms in this group to achieve the 40% total acreage for the county. The same exercise was conducted to represent sorghum at a rate of 10% to achieve the county mean of 12%. Finally, the acres allocated were tallied and subtracted from the total to represent wheat grown by the farms.

Once the farming sizes and commodities grown were simulated, a number of precision agriculture parameters were applied to the group. There are five groups of equipment and practices divided among the farms: yield monitors (40%), auto guidance (20%), variable rate technology (10%), big data analytics (10%), and telematics (5%) (Schimmelpfennig and Ebel 2011).

Next, yield values were calculated based on a normal distribution with the mean value being the five-year average for each representative commodity. Based on USDA NASS data, mean yield and standard deviation were assigned each crop (Table X). Corn has an average yield of 150 and standard deviation of 50, soybeans has an average yield of 40 and standard deviation of 15, sorghum has an average yield of 95 and a standard deviation of 40, and wheat has an average yield of 45 and a standard deviation of 10. With the base yield established, precision agriculture variables increased the yield of the commodities based on the equipment present within the operation. Farms utilizing variable rate technology increased yield by 15% using a normal distribution. Farms with an auto guidance system had a yield increase of 20% applied, also on a normal distribution. Finally, for farms using analytics, which requires a yield monitor to be used, an increase of 15% was added to yield based on a normal distribution (Schimmelpfennig and Ebel 2011).

Using current market data regarding inputs and commodity prices within Clay County, a net farm income was calculated for each farm based on the commodities grown, the inputs, the PA attributes, and the yield. Credits were issued to the cost of inputs as a means of simulating the reduction of input usage based on the PA technology adoption of each farm. A grower with an auto guidance system had a 5% decrease in input costs,

analytics had a 15% decrease, and telematics decreased costs by 10%, all based on a normal distribution.

Input prices are based upon a Clay County, Kansas agricultural input retailer's price list using average prices for seed varieties grown locally and fertilizer and herbicide prices. The recommendations are based upon common recommendations and practices for the area (Table 4.5). To bring some reality to the agronomic decisions built within this simulation, it is assumed within the herbicide program for example, that the grower will have at least one burndown application either following harvest or prior to plant of the next year's crop. The commodity prices for sales of crops were based upon the March 13, 2015 harvest-time closing prices for AgMark, LLC elevators in Clay Center, KS.

4.3 Development of Dataset in R

The entire project was developed and produced using an analytics package written in the computational language of R (R Core Team, 2015). Per the introductory paragraph on The R Foundation's, or R-Project's webpage: "R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows, and MacOS" (2015). R represents an "integrated suite of software facilities for data manipulation, calculation and graphical display." Essentially, R is the language and physical environment that exists for the purpose of calculating, computing, and manipulating data. The platform gives the statistician and economist the ability to input data and run many different packages to manipulate and analyze the data set in the forms of "linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering as well as graphing" (The R Foundation 2015). The following will step the novice user of R through the thought process and script used during the inception of this dataset and analysis.

As noted earlier, because the dataset was not directly observed, all of the data was regenerated using simulation techniques found in R. The development of this data was based upon known values and represents an environment comparable to reality. The reasoning behind using simulated values is that the private values comprising the actual dataset are proprietary to the owner-members of the cooperative. The design, however, follows a similar format and pattern to the actual, observed data.

To begin the generation of the dataset, a histogram representing Clay County Kansas in the 2012 agricultural survey (Figure 4.1) was selected as the basic framework for determining farm size. The following code was used to create a dataset simulating this distribution:

```
rtruncnorm(n = 20, a = 1, b = 9, mean = 1, sd = 50)
```

where “rtruncnorm” represents the command to simulate a truncated normal distribution in the first group of the histogram, 20 growers in total, ranging from acreage of 1 to 9, and a standard deviation of 50.

The next step in the process was to divide the total acres created within the dataset into four commodities: corn, soybeans, sorghum, and wheat (Table 4.1, Table 4.2, Table 4.3). To do this, the acres were allocated on a normal distribution according to the following parameters:

- Corn was be grown by farms with 1000 acres or more of total farmland.
- Soybeans were grown on all farms. Farms larger than 1000 acres allocated 43% of acres to planting soybeans, while farms smaller than 1000 acres allocated 1/3.
- Sorghum was also grown on all farms, allocating 1/3 of total farmland.

- Wheat acres were allocated at a value of the total acreage minus the values of the three previously listed commodities. As a result, some farms did not have wheat acres.

Table 4.1: Distribution of Acres as a Percentage of Total Simulated Acres by Crop

Farm Size	Corn	Soybeans	Sorghum	Wheat
<1000 Acres	0%	26%	84%	30%
>1000 Acres	100%	74%	16%	70%

Table 4.2: Distribution of Acres as a Percentage of Total Acres by Farm Size

Farm Size	Corn	Soybeans	Sorghum	Wheat
<1000 Acres	0%	33%	33%	34%
>1000 Acres	18%	43%	3%	36%

Table 4.3: Distribution of Acres by Crop (Simulated Values)

Farm Size	Corn	Soybeans	Sorghum	Wheat	Total
<1000 Acres	0	38,096	38,096	38,211	114,403
>1000 Acres	45,400	106,710	7,385	89,135	248,630
Total	45,400	144,806	45,481	127,346	363,033

An example of this type of command would look like the following:

```
ifelse(dat.sim$Total>999, abs(rnorm(541)*0.215*dat.sim$Total),0)
ifelse(dat.sim$Total >999, dat.sim$Total*.43, dat.sim$Total*0.333)
```

where the dataset is generated using an if/then statement for corn, soybeans, and sorghum and then subtracting those three values from the total for wheat.

The next task at hand was to simulate the PA attributes to represent equipment purchases and adoption of technology on the farm. Similar to reality, only farms within the model with 500 or more acres have the ability to own a yield monitor. Likewise, only farms of the same size, 500 or more acres, have access to an auto guidance system. All farms have the ability to have variable rate technology due to the ability to have products custom-applied. The most limited resource deals with access to big data analytics, which require the farms to have both a yield monitor and auto guidance to qualify. Similarly,

telematics would only be a necessary option to have on the farm if the operator was using big data analytics, thus requiring the farm to have the same criteria as big data to qualify for access to the technology.

To simulate the adoption of precision agriculture technology, the variables were added to the dataset in the form of a binomial response - 1 for yes, 0 for no. A sample command would look like the following:

```
ifelse(BIGDATA==1, rbinom(541, 1, .155),0)
aggregate(Total~TELEMATICS, data=dataset, sum)
sum(Total Acres)
aggregate(Total~TELEMATICS, data=dataset, sum)[2,2]/sum(Total Acres)
```

where the attribute is allocated on a binomial distribution with a probability of 15.5% to satisfy the percentage of acres represented by this application.

Once all of the farm's attributes were created, using the information from the previous section, each commodity's yield was generated on a normal distribution where the mean was representative of the five-year average of the county crop yield and the standard deviation, similarly being based on the five-year minimum and maximum average yield. A sample command is listed below:

```
SB Acres<-rnorm(541, Mean=40, Standard Deviation=15)
SB Yield<-abs(SB Acres)
SB VRT<-ifelse(dat.sim$VRT==1, SB Yield*rnorm(1, Mean=0.15, Standard
Deviation=0.25),0)
SB Auto Guidance<-ifelse(AUTOGUIDANCE==1, SB Yield*rnorm(1, Mean=0.2,
Standard Deviation=0.25),0)
ifelse(Big Data Variable==1, SB Yield*rnorm(1, Mean=0.15, Standard
Deviation=0.25),0)
Soybeans Yield<-absolute value(SB Yield + VRT + Auto Guidance + Big Data)
```

where the simulated yield is a function of a random normal distribution where the county average yield of soybean over the five-year period was 40 bushels per acre and standard deviation was 15 (statistics for all crops are shown in Table 4.4). The yield was made an

absolute value because, while numerically and statistically that would be possible in a normal distribution, negative yield is not possible in reality. After that, a credit was applied based on the expected yield gain as a result of the implementation of the different PA variables. Each farm with a positive value returned within the binomial distribution had a yield bump of 15% for having variable rate technology (VRT), 20% for having an auto guidance system (AG), and 15% for using software for data analytics (BIGDATA). The percentages applied based on technology was based, again, on a normal distribution.

Table 4.4: North Central Kansas Average Yield 2005-2014

Commodity	N	Mean	Std Dev	Minimum	Total
Corn	10	121.31	18.24	89.20	150.00
Soybeans	10	37.84	6.28	24.90	47.00
Sorghum	10	88.10	15.08	55.40	108.00
Wheat	10	40.85	5.32	31.10	47.00

Source: (National Agricultural Statistics Service 2015)

The final piece to be simulated was input costs. Because this study focuses on technical efficiency as a result of implementing precision agriculture technology, prices are for the most part, irrelevant. Thus, prices will be equal for both the inputs and grain sales of all farms. Thus, the prices of inputs and outputs will not be spread on a normal distribution, but rather, allocated according to the number of acres for inputs and number of bushels for outputs.

Table 4.5: Input Costs by Commodity

Commodity	Seed	Fertilizer	Herbicide	Total
Corn	\$90.00	\$149.00	\$59.05	\$298.05
Soybeans	\$65.00	\$20.00	\$73.54	\$158.54
Sorghum	\$23.40	\$104.50	\$59.05	\$186.95
Wheat	\$19.20	\$76.50	\$36.91	\$132.61

The script used for the application of input costs and gross revenue is shown below:

Input Costs:

```
Sorghum Input Costs<-ifelse(GS Acres >0.1, (GS Acres*186.95),0)

Auto Guidance Input Savings<-ifelse(AUTOGUIDANCE==1,GS Input*rnorm(1,
Mean=0.05, Standard Deviation=0.25),0)

Big Data Input Savings<-ifelse(BIGDATA==1, GS Input*rnorm(1, Mean=0.15,
Standard Deviation=0.25),0)

Sorghum Input Costs<-absolute value(GS Input-Auto Guidance-Big Data-
Telematics)

Telematics Input Savings<-ifelse(TELEMATICS==1, GS Input*rnorm(1,
Mean=0.1, Standard Deviation=0.25),0)
```

Gross Revenue:

```
Sorghum Revenue<-ifelse(Sorghum Acres > 0.1, (Sorghum Yield*Sorghum
Acres*3.95),0)
```

where, again, the input costs are the same for all firms for each commodity (Table 4.5).

Similar to the increase in yield resulting from the adoption of precision technology, the value of inputs decrease in the same manner based on the inclusion of each variable on the farm. Auto guidance decreased input costs by 5% and data analytics decreased input costs by 15%. The allocation of the input cost savings was done so on a normal distribution with the mean value for each matching values found in Table 4.7. For revenue, the value was determined via a simple multiplication of the bushels raised per farm and the current 2015 fall crop elevator price in Clay Center, Kansas (Table 4.6).

Table 4.6: Clay Center, KS 2015 New Crop Commodity Prices (03/13/2015)

Commodity	Futures Month	CBOT Price	Basis	Cash Price
Corn	DEC 2015	\$4.05	-0.45	\$3.60
Soybeans	NOV 2015	\$9.53	-0.80	\$8.73
Sorghum	DEC 2015	\$4.05	-0.10	\$3.95
Wheat	JUL 2015	\$5.44	-0.42	\$5.02

Source: (AgMark, LLC 2015)

Table 4.7: Influence of PA Technology Adoption on Input Cost and Yield

PA Technology Adopted	Yield Increase	Input Cost Decrease
Yield Monitor	0%	0%
Variable Rate Technology (VRT)	15%	0%
Auto Guidance	20%	5%
Big Data Analysis	15%	15%
Telematics	0%	10%

Once all of the variables were allocated, bushels simulated, and inputs and outputs were determined, the total Net Farm Income (NFI) was calculated by adding each commodity's net return. From there, the net return was divided by the total acres farmed to achieve the per-acre value of net return for the farm.

4.4 Analytical Approach

First and foremost, it is important to note that the purpose of this analysis was to simulate a benchmarking scenario where observed empirical data could easily be inserted within the model to generate a real response as to the grouping of peers, measurement of technical efficiency, and basis of comparison for future surveys. As noted, the dataset being created will be as close to real as is possible given the constraint of public information and availability of survey results. The hypothetical dataset generated will then be subjected to the same analysis as empirical data would be, and the cherry-picking of the top entries will be studied in a similar fashion to explore differences between other farms

within assigned peer groups. When real data does become available, the variables of interest and hypotheses outlining expected outcomes will already be identified for further study and exploration within the confines of this work.

The analysis was similar to that of Curtiss and Jelinek in their study of cost efficiency on Czech wheat farms. Their usage of NFI as a means to analyze the efficiency scores in determining how precision farming variable affect the profitability of the farm provides an important keystone to the study presented here. As a means to measure efficiency, all levels of output were brought to a dollar valuation based upon economic data. Then, all firms were collectively analyzed through DEA to determine the impact that each level of technological adoption had upon profitability in comparison to farms of similar attributes.

Their research found that, on average, farms which adopted varying levels of precision technology carried the potential to reduce on-farm costs by as much as 37%. Their results found that “lower levels of allocative efficiency compared to technical efficiency scores imply that there is a greater potential for decreasing costs through correcting for input combinations (allocation) through different production practices (technologies) than in the proportional adjustment of input levels captured by technical efficiency” (Curtiss and Jelinek 2012). Essentially, they determined that the greatest potential to decrease costs exists when inputs are properly allocated through the application of a precision technology as a means of determining the proper rate and placement. Curtiss and Jelinek’s research also showed varying levels of efficiency scores, suggesting that farms identifying as adopters of PA technology had the potential to reach greater economic returns than their counterparts (2012).

To replicate a study similar to this, the data envelopment to be conducted will compare the NFI to the total acreage, or basically the output, being a measurement of the net returns to the farm based upon the resources of acres. This will then conclude which farms are able to return the most profit to the farm based upon the acres they have readily available. Much like Curtiss and Jelinek's study, all measures of input and all measures of output have been converted to the same numerical format – in this case – U.S. dollars to be compared as a per unit, per acre, or per farm value. Because all farms have the same methodology of deriving at a measurement of profitability, it makes the math far simpler to calculate and compare. This would not necessarily have to be the case though. Even if some of the firms had allocated acres to livestock production, or perhaps, left some of the ground fallow, this type of analysis will allow for the comparison of all operations regardless of their true allocation of acreage and/or inputs because efficiency will be measured in comparison of peers and the ability to convert acreage to NFI.

4.5 Benchmarking in R

The DEA was conducted using the package *Benchmarking* (Bogetoft and Otto 2014) found in R. This package was chosen over the *Nonparaeff* (Oh and Suh 2013) analysis package primarily on the basis of personal preference, but also due to the increased capabilities to conduct scale efficiency analysis through DEA.

Upon loading the *Benchmarking* package, the X and Y coordinates were identified where the X-axis represented the Total acres represented by each farm and the Y-axis represented the NFI. At that point, the scale efficiency of the dataset was calculated using the following logic:

```
dea(X,Y,ORIENTATION="input",Returns To Scale ="vrs")
```

where “**dea**” represents the scale efficiency DEA command as a function of the X and Y identified previously and “**vrs**” represents the assumption of varying returns to scale exists. The same function was calculated once more assuming constant returns to scale. For the calculation of scale efficiency, the “**crs**” value will be divided by the “**vrs**” value.

This calculation of scale efficiency represents the primary motive for the study. It is believed that this number, in combination with the varying levels of adoption of PA technology, will demonstrate the effectiveness of technology and profitability potential which may exist on the farm.

CHAPTER V: DATA AND RESULTS

5.1 Data Introduction

This section describes the observed results of the data envelopment analysis (Appendix B) conducted upon the dataset generated from the script outlined in Chapter 4 (Appendix A). The hypothesis going into the analysis was that the farms with the greatest adoption of precision agriculture technology would have the greatest efficiency levels. The following will outline and interpret the observed results, demonstrating the economic impact which can be inferred upon a dataset of empirical values.

5.2 Descriptive Statistics of Simulated Values

The simulated dataset generated 541 individual farms ranging in size from 1 acre to 4,181 (Table 5.1). Some observations of the dataset were that not all farms recorded a positive net farm income (NFI), and as a result, those farms will not be represented in our analysis. In all, there were 19 farms out of 541 which did not have a positive return, with none of the farms within this group having adopted any level of PA technology. The largest grower with a negative NFI had a farm size of 1832 acres and was the only farm in this group to both produce corn and be of a size over 1000 acres. The negative return for this farm was attributed to the low soybean yield slightly above 5 bushels per acre, generating a negative net return on this commodity of over \$86,000. In reality, it would be assumed that the negative return would be offset by an insurance payment in support of the low yield return. Because that information does not exist within the confines of this dataset, the only correction to this observation will come when real data replaces the simulation.

Overall, the dataset generated is very similar to the observed values of this county. There are some factors not being taken into account in this – namely crop sharing and cash

rent, which could increase farm size and farm income considerably – but when evaluating the farms as they were generated, the concern will be the return of the farm against the acres of the farm, which would be the same methodology regardless of the structure of the business. There were some outliers identified as well, especially noted when yield values were well below average with no correction based upon crop insurance or subsidy programs. While these values are seen in reality, usually a profit correction is made at the enterprise level to correct the below average yield and offset the negative NFI.

Again, because this model is to generate net farm income across many firms, the end result will be to compare farm at the values simulated, regardless of the true-to-life implications that low yields will have on profitability and the risk management practices that assist in these situations. The correction for this will be replacing the simulated crop income values with observed data.

Table 5.1: Descriptive Statistics (Simulated Values)

Variable	N	Mean	Std Dev	Minimum	Maximum
Total Farm Acres	541	670.18	801.15	1.00	4181.00
Corn Acres	131	346.56	303.54	5.37	1819.27
Soybean Acres	541	267.66	349.30	0.33	1797.83
Sorghum Acres	541	84.07	78.68	0.26	329.67
Wheat Acres	538	236.70	316.02	0.33	1745.61
Corn Yield	131	134.97	27.52	89.89	243.93
Soybean Yield	541	42.00	21.23	5.01	99.64
Sorghum Yield	541	80.37	37.15	5.46	149.57
Wheat Yield	538	44.94	21.03	5.00	134.97
Corn Inputs	131	280.31	39.60	174.83	298.05
Soybean Inputs	541	152.12	21.06	95.49	158.54
Sorghum Inputs	541	187.16	7.24	71.34	227.22
Wheat Inputs	538	128.71	15.87	83.12	132.61
Efficiency	541	0.92	0.21	0.00	1.00

When comparing across all firms, a linear regression (Table 5.2) demonstrates a best-fit line measuring net farm income per acre based upon the four precision equipment

adoption explanatory variables plus total farm acres. The regression coefficients tested whether the data were representative of the characteristics of local farms and their calculation of Net Farm Income. The regression output registers only VRT variables to be statistically significant at a level less than 1%, two other variables (Auto Guidance and Telematics) to be statistically significant at the 5% level, and a standard error of 83.73 based upon 534 degrees of freedom. One item in particular that did stand out in the regression results was that ownership of a Yield Monitor had a negative impact on NFI. When the simulated dataset is replaced with actual observations, the OLS regression may reflect a completely different correlation. Given the acceptable signs and magnitudes of regression coefficients, the dataset was deemed acceptable for further analysis via DEA.

Table 5.2: OLS Regression of Simulated Dataset

Coefficients	Estimate	Std. Error	t-value	P> t 	Significance
(Intercept)	138.90	4.78	29.040	0.0000	<1%
Yield Monitor	-2.68	12.48	-0.22	0.8301	
VRT	58.72	13.35	4.40	0.0000	<1%
Auto Guidance	29.67	16.64	1.78	0.0751	<5%
Big Data	2.79	31.66	0.09	0.9298	
Telematics	60.54	14.53	-0.18	0.8590	<5%
Total Acres	0.01	0.01	1.33	0.1834	
Adjusted R ² :	0.0644				
p-value:	2.12 x 10 ⁻				
Std Error:	83.73				
Degrees of	534				

5.3 Scale Efficiency Analysis of Simulated Dataset

As noted in Chapter 4, the goal of this project would be to determine the technical efficiency of farms based on the two variables of Net Farm Income and Acreage in a date envelopment of variable returns to scale versus constant returns to scale. The mathematics in this process calculated the scale efficiency of the farms in the study. From there, a scatter plot was formed (Figure 5.1) to visually diagram the collection of farms and their positioning along or below the efficient frontier. The firms represented that sit on the frontier directly received the highest score possible in this study of 1.0 out of 1.0. Likewise, as noted by the summary statistics in Table 5.3, the mean scale efficiency score for all farms in the model was 0.92, representing the potential to increase efficiency on average by 8%. Any farm with a negative value for Net Farm Income scored a zero for scale efficiency. An interesting note is that no farm which adopted any form of technology recorded a negative Net Farm Income (noted by the fact that the minimum value for all PA technology attributes is greater than zero).

The second regression which was run was the Censored Tobit Regression Model (Table 5.4). This model represents the expected increase in scale efficiency as technology is adopted. The intercept is the base number upon which the score is calculated. Because some purchases required the adoption of multiple pieces of technology, the value would be based upon the highest level of adoption. Because two of the primary precision agriculture explanatory variables were not recorded as significant, there may be multiple concerns with using these variables to describe efficiency gains on the farm. In reality, big data analysis and telematics are two pieces of technology that hold incredible value in the statistical sense, but do not always provide a direct return in the ways that the other technologies do

to the farmer. Again, until the model is re-run using actual observations, this is simply an analysis of the data presented.

Figure 5.1: DEA Plot – Scale Efficiency of All Farms

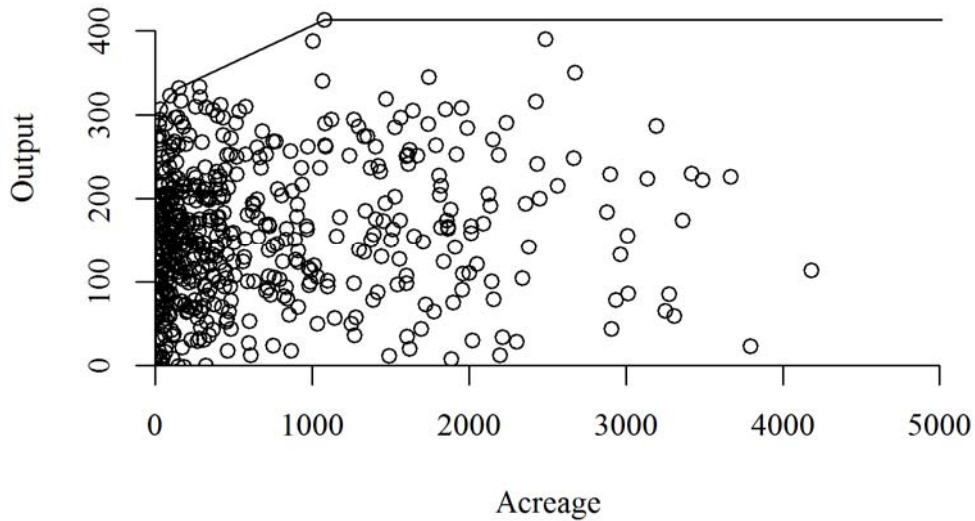


Table 5.3: Summary Statistics for Scale Efficiency Scores

	Mean	Std Dev	Min	Max
Yield Monitor	0.72	0.26	0.04	1.00
VRT	0.79	0.22	0.09	1.00
Auto Guidance	0.78	0.25	0.14	1.00
Big Data	0.76	0.18	0.34	1.00
Telematics	0.80	0.19	0.46	1.00
All Farms	0.92	0.21	0.00	1.00

Table 5.4: Censored Tobit Regression Model

Coefficients	Estimate	Std. Error	t-value	P> t 	Significance
(Intercept)	0.90	0.01	85.52	0.0000	<1%
Yield Monitor	0.08	0.03	3.06	0.0023	<1%
Auto Guidance	0.09	0.04	2.12	0.0341	<1%
VRT	0.07	0.03	1.96	0.0502	<5%
Big Data	-0.08	0.08	-0.99	0.3221	
Telematics	-0.02	0.09	-0.21	0.8323	
logSigma	-1.56	0.03	-49.48	0.0000	<1%

5.4 Scale Efficiency DEA for Commodity Sub-groups

When conducting a data envelopment analysis, it is possible to examine the firms within the study based upon attributes as well. When the analysis is re-run to account only for the attribute, the efficiency scoring is re-worked to properly benchmark within the new group of firms. In this case, the analysis was conducted once more for each of the four commodity groups. The technical efficiency overall remains the same, while the scoring within sub-group adjusts according to the highest ranking farms, where a lower scoring firm, in essence has the capability of being considered inefficient in one group to perfectly efficient within the second group. For example, there are two efficient farms on the frontier on the Corn Producers DEA Plot (Figure 5.2). The second producer on the plot, lined directly below the top-producing farm, is in the exact same position of the plot (Figures 5.3, 5.4, and 5.5) for the other three commodities represented in the study. The second producer may be considered efficient among the corn producers, but is no longer

represented on the frontier for the other three commodities. This same analysis holds true with regards to this farm's performance on the efficiency frontier gauging all farms.

Figure 5.2: DEA Plot – Corn Producers

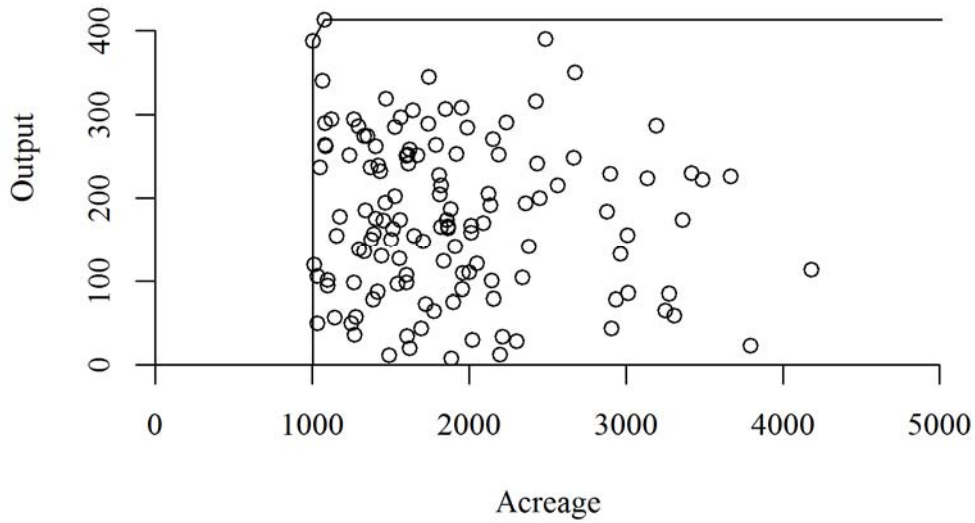


Figure 5.3: DEA Plot – Soybean Producers

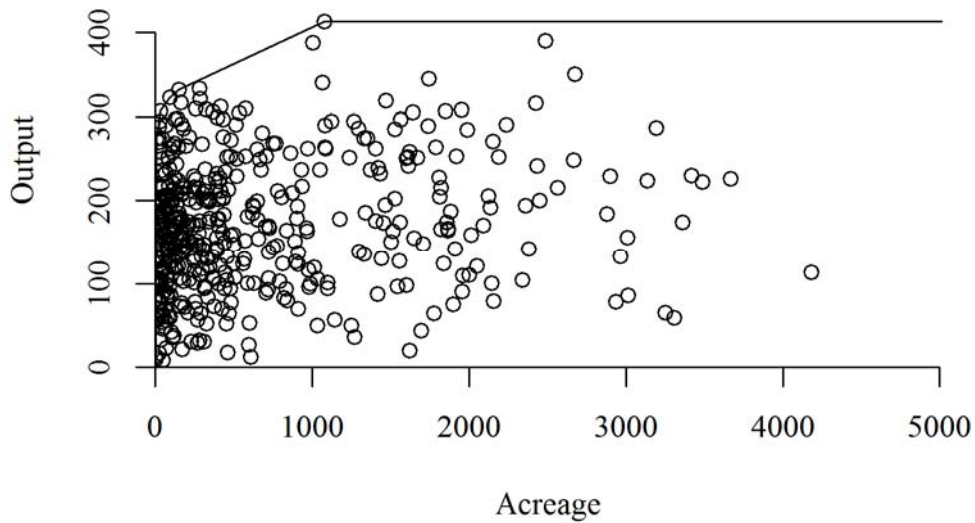


Figure 5.4: DEA Plot – Grain Sorghum Producers

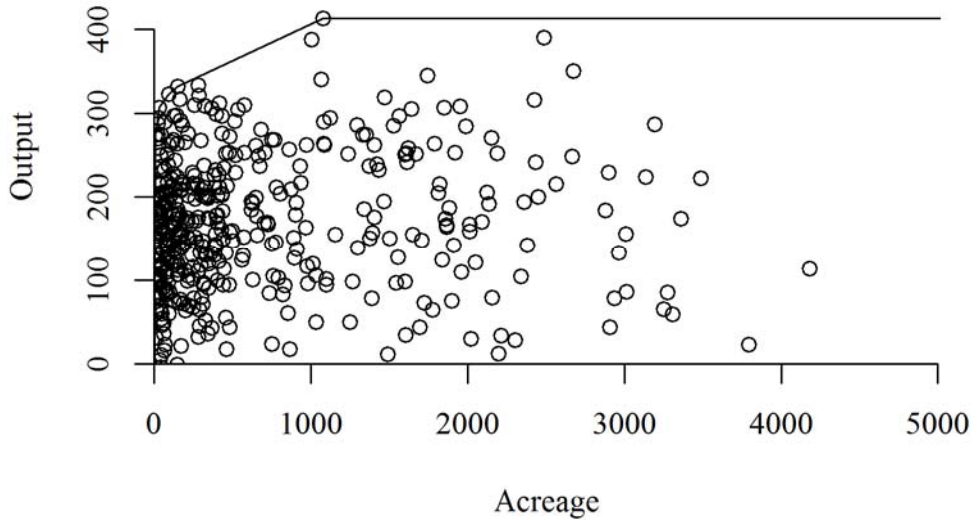
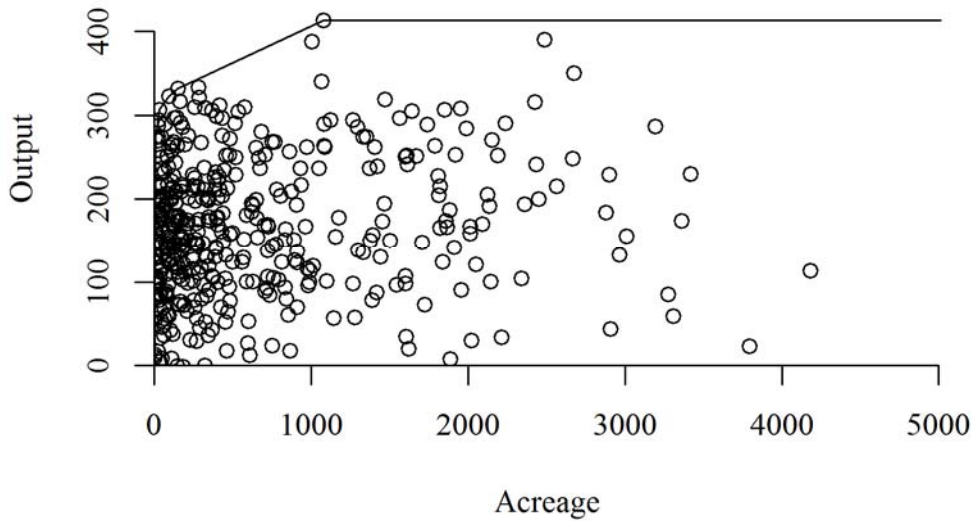


Figure 5.5: DEA Plot – Wheat Producers



5.5 Scale Efficiency DEA for Precision Agriculture Technology Sub-groups

As noted in the previous section, the efficiency calculation was conducted once for all farms in the study and then multiple data envelopment analyses were conducted based on the sub-groups of interest. This section notes some of the highlights found in the DEA of the precision agriculture technology sub-groups.

Also interesting to note, of the ten adopters of the highest level of technology available in this study (telematics), the lowest scale efficiency was at a level of 0.94 when

compared to the entire group of farms. Three of the ten were of a size 1000 acres or less, and only one grower had adopted all levels of technology (the other firms had not adopted variable rate technology). This grower was the largest farm of the group, had adopted all phases of technology, and yet had the lowest efficiency score. This farm had yielded well above the average on both corn and wheat, but had yielded well below the average on soybeans and sorghum. In a dataset of real, not simulated, observations, this would probably be the most interesting farm to follow due to its size and technology adoption. If, perhaps, the loss was attributed to weather or some other disaster, it would be expected that other farms would have suffered similarly. Because this study did not include geographic area, farm attributes including irrigation, and other details to allow for further examination, it would be difficult to offer a sincere opinion on how to improve based solely on the given information.

Figure 5.6: DEA Plot – Adoption of Yield Monitors

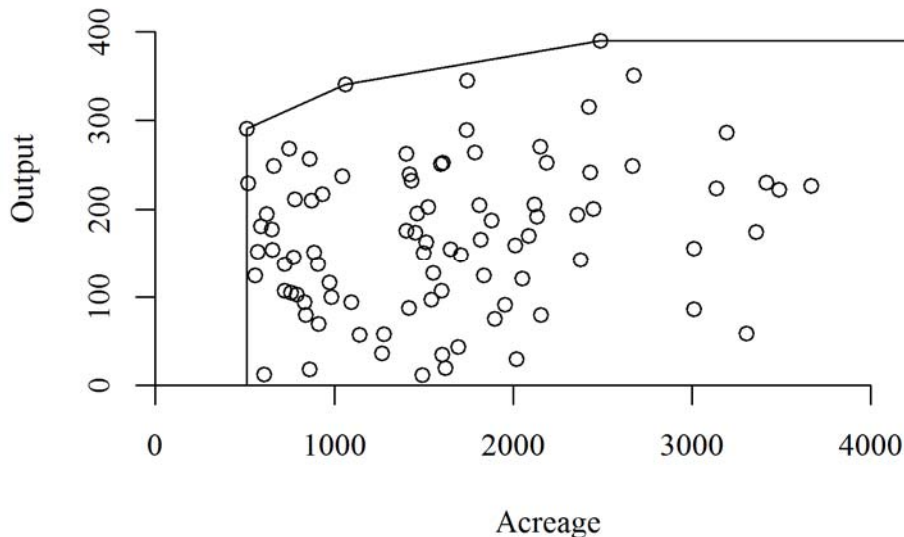


Figure 5.7: DEA Plot – Adoption of Variable Rate Technology

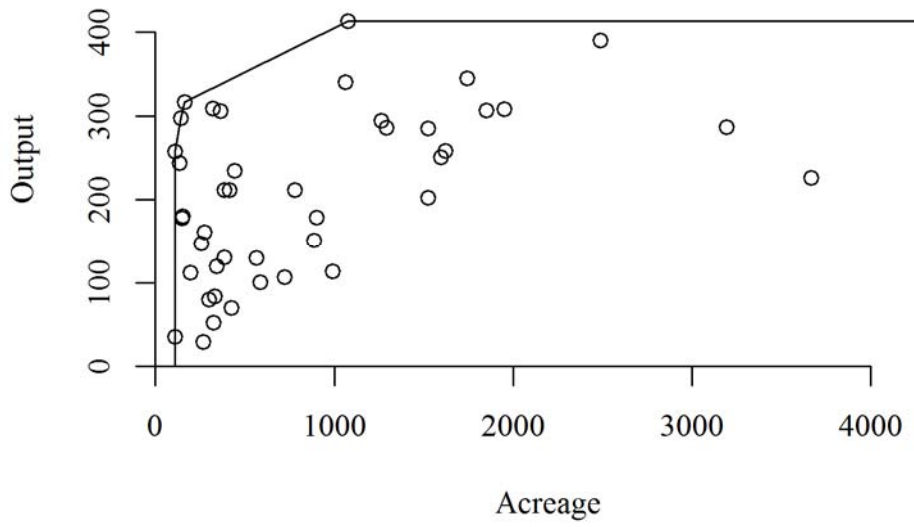


Figure 5.8: DEA Plot – Adoption of Auto Guidance

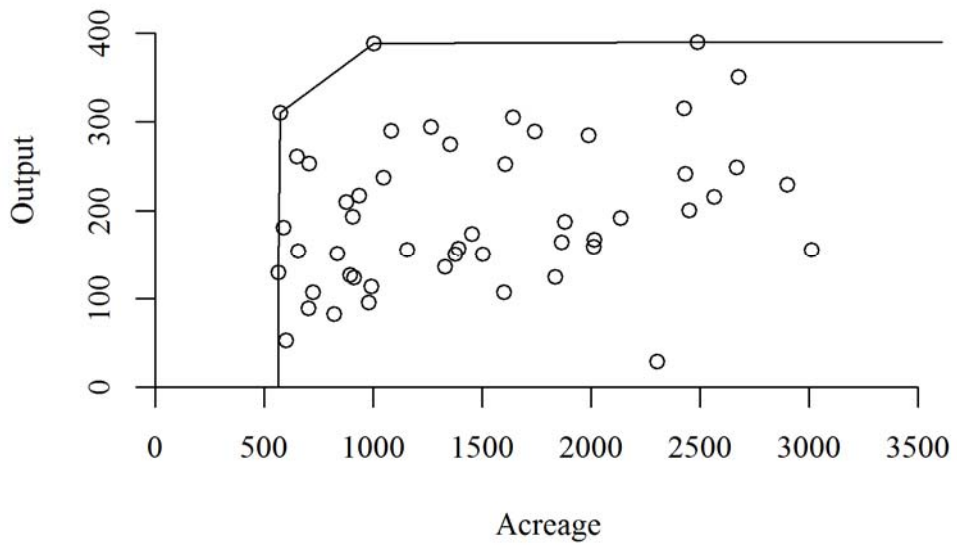


Figure 5.9: DEA Plot – Adoption of Big Data Analytics

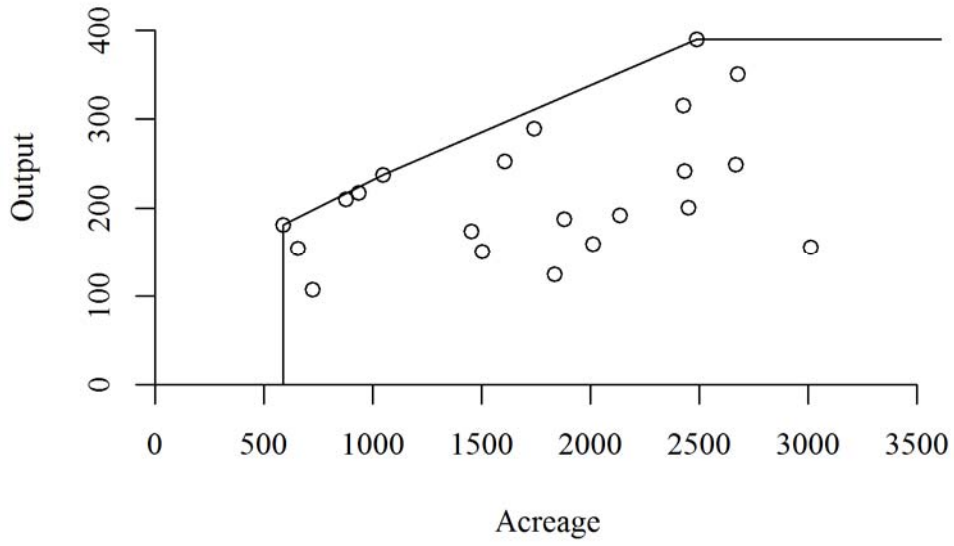
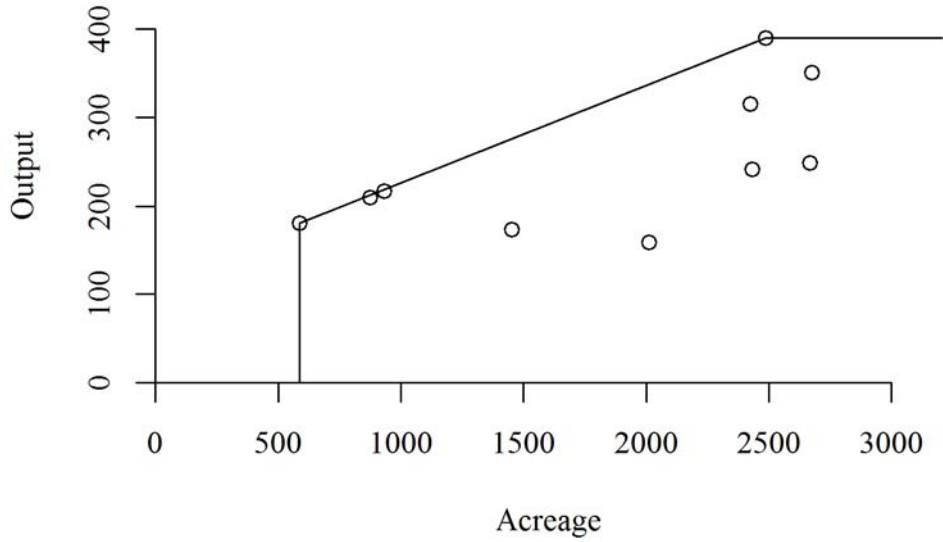


Figure 5.10: DEA Plot – Adoption of Telematics



CHAPTER VI: CONCLUSION

The purpose of this study was to create a platform for both calculating technical efficiency as well as benchmarking the farms against their peers as a means of potentially outlining areas in which the farms could improve. This study used a simulated dataset that was expanded based upon real observations, but in the end, the results were an example of the type of economic analysis this dataset allows for. The goal is to use the R script contained in Appendix B to conduct an analysis of real observed data.

Data envelopment analysis has the potential for addressing many farm-level issues by benchmarking across a community of peers. A given farmer's technical efficiency can be directly compared to their peer group to determine if the farm is making the best use of input relative to output.

These methods have the potential to replace informal 'coffee shop talk' and rumor with quantitative results suitable for making adjustments to farming operations. Techniques such as the DEA adapted to cooperative members are likely to impact the agricultural industry by improving big data analytics.

The results of the analysis conducted upon the simulated dataset revealed several different challenges that will have to be overcome in order for this type of strategy to work in determining efficiency and outlining ways to improve. Some of the crucial variables that must be taken into consideration include how the dataset handles losses due to weather or other natural events, how to evaluate a dryland versus irrigated farming operation, how to deal with crop sharing and cash rent, and how to offer suggestions for improvement when all levels of technology have been adopted.

While this study in the end offered more questions than answers, the purpose of this work was to do just that – to begin the conversation of how to properly evaluate the

effectiveness of technology and how to value the potential losses or gains as a result of adopting, or not adopting, precision agriculture on the farm. Benchmarking using DEA appears to be capable of providing a proper stepping stone in beginning to evaluate precision agriculture technology. The next phase of the study, implementation of the model using observed data, will give a much clearer impression of whether or not this process will meet the goals set.

6.1 Future Work

This study has inspired a number of future projects as well as a number of additions that will be required to increase the effectiveness of what is presented here for usage in industry. As outlined in a few of the sections within this study, there will be a few explanatory variables required to be added in order to assist in increasing how effective the existing script will work on. These additions will not be limited to the following: irrigated versus non-irrigated ground, owned versus leased acres, pasture versus crop ground, and year of adoption for the different levels of precision technology. The addition of these new variables will help to better evaluate the efficiency of the farms taking part in the survey. Other items that have been considered for further evaluation in the study include a variable for rainfall as well as a better way to account for losses due to insect infestations, disease, or other natural events. This inclusion will assist in being able to make recommendations based on the realized conditions present rather than on statistical analysis alone.

While this study represents a way to evaluate a mass quantity of farms with a single analysis conducted, the real analysis will take place on an individual farm-by-farm basis. This study brought a brand new concept in evaluating technical efficiency amongst farms, but also inspired a brand new study which will begin upon completion and implementation of this project. The new concept will conduct a similar DEA. This time, instead of the

DMUs being separate entities, the DMUs will be sub-field areas. A grid will be superimposed upon a field in a manner similar to Figure 6.1 with the results of the scale efficiency study and DEA generating a visual analysis of technical efficiency on a by-acre basis.

Figure 6.1: Field Technical Efficiency

75	85	86	92
64	55	95	90
66	78	99	91
88	87	50	88
95	91	90	89

This will be conducted using as-applied maps of products not applied at a uniform rate as a means of determining input and yield maps as a means of determining output. The thought being that multiple agronomic layers could be laid upon this diagram to determine the economic results of applying products or making decisions for future applications of inputs. Given the geo-referenced technical efficiency scores, tests for spatial autocorrelation can determine whether productivity remains spatially clustered, thereby providing insights into natural or man-made variability. Spatial autocorrelation indicates how similar the values of a variable are with respect to distance.

REFERENCES AND WORKS CITED

- AgMark, LLC. *www.agmarkllc.com*. March 13, 2015.
<http://www.agmarkllc.com/cashprices> (accessed March 15, 2015).
- Berg, S. *Water Utility Benchmarking: Measurement, Methodologies, and Performance Incentives*. Edited by Maria Luisa Corton et al. London: IWA Publishing, 2010.
- Bogetoft, Peter, and Lars Otto. *Benchmarking with DEA, SFA, and R*. New York: Springer Science, 2011.
- . "Benchmarking: Benchmark and Frontier Analysis Using DEA and SFA." *R package version 0.24*. 2014.
- Burkett, John. "Notes on choosing DEA models, writing R code, and interpreting R output." September 26, 2011.
- Curtiss, Jarmila, and Ladislav Jelinek. "Cost Efficiency and Farm Self-selection in Precision Farming: The Case of Czech Wheat Production." *EAAE Seminar 'Innovation for Agricultural Competitiveness and Sustainability of Rural Areas'*. Prague, Czech Republic, 2012. 19.
- Farrell, M J. "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society Series A (General)*, vol. 120, no. 3 (1957): 253-290.
- Featherstone, Allen M, and Md. Habibur Rahman. "Nonparametric Analysis of the Optimizing Behavior of Midwestern Cooperatives." *Review of Agricultural Economics* 18, no. 2 (May 1996): 265-273.
- Featherstone, Allen M, and Terry L Kastens. "Non-Parametric and Semi-Parametric Techniques for Modeling and Simulating Correlated, Non-Normal Price and Yield Distributions: Applications to Risk Analysis in Kansas Agriculture." *Journal of Agricultural and Applied Economics*, August 2000: 267-281.
- Featherstone, Allen M, Michael R Langemeier, and Mohammad Ismet. "A Nonparametric Analysis of Efficiency for a Sample of Kansas Beef Cow Farms." *Journal of Agricultural and Applied Economics*, July 1997: 175-184.
- Fleming, Euan, Terence Farrell, Renato Villano, and Pauline Fleming. "Is Farm Benchmarking the New Acceptable Face of Comparative Analysis?" *Australasian Agribusiness Review* 14, no. 12 (2006): 18.
- Griffin, Terry W. *Decision-Making from On-Farm Experiments: Spatial Analysis of Precision Agriculture Data*. PhD Thesis, West Lafayette, IN: Purdue University, 2006.
- Griffin, Terry W, Craig L Dobbins, Tony J Vyn, Raymond J G M Florax, and James M Lowenberg-DeBoer. "Spatial Analysis of Yield Monitor Data: Case Studies of On-

- Farm Trials and Farm Management Decision Making." *Precision Agric*, 9 2008: 269-283.
- Griffin, Terry, J Lowenberg-DeBoer, D M Lambert, J Peone, T Payne, and S G Deberkow. *Adoption, Profitability, and Making Better Use of Precision Farming Data*. Staff Paper #04-06, West Lafayette, Indiana: Department of Agricultural Economics, Purdue University, 2004.
- Groover, Gordon, and Robert Grisso. *Investing in GPS Guidance Systems?* Extension Publication 448-076, Blacksburg, Virginia: Virginia Cooperative Extension, Virginia Tech University, 2009.
- Henningsen, Arne. "censReg: Censored Regression (Tobit) Models." *R package version 0.5-20*. 2013.
- Kutter, T, S Tiemann, R Siebert, and S Fountas. "The Role of Communication and Cooperation in the Adoption of Precision Farming." *Precision Agric*, 12 2011: 2-17.
- Mark, Tyler B, Brian Whitacre, and Terry W Griffin. "Assessing the Value of Broadband Connectivity for Big Data and Telematics: Technical Efficiency." *Southern Agricultural Economics Association's 2015 Annual Meeting*. Atlanta, Georgia, 2015.
- Moghnieh, Ghassan A, Allen M Featherstone, and Barry K Goodwin. "A Nonparametric Analysis of Cost Minimization and Profit Maximization Behavior for a Sample of Kansas Farms." Staff Paper, Department of Agricultural Economics, Kansas State University, Manhattan, Kansas, 1991, 15.
- National Agricultural Statistics Service. *USDA/NASS QuickStats Ad-hoc Query: 2005-2014 North Central Kansas Commodity-Specified Yield* . March 25, 2015.
- National Agricultural Statistics Service. *USDA/NASS QuickStats Ad-hoc Query: 2005-2014 Survey Clay County Kansas Grain Acres Harvested*. 2015.
- National Agricultural Statistics Service. *USDA/NASS QuickStats Ad-hoc Query: 2007 Census Clay County Kansas Internet Access Total*. n.d.
- Oh, Dong-hyun, and Dukrok Suh. "nonparaeff: Nonparametric Methods for Measuring Efficiency and Productivity." *R package version 0.5-8*. 2013.
- Oliver, M A. *Geostatistical Applications for Precision Agriculture*. New York: Springer Science, 2010.
- Parman, Bryon James. *Economies of Scale for Data Envelopment Analysis with a Kansas Farm Application*. PhD Thesis, Manhattan, KS: Kansas State University, 2013.
- Pindyck, Robert S, and Daniel L Rubinfeld. *Microeconomics*. Sixth Edition. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

- Quintana-Ashwell, Nicolas E, and Allen M Featherstone. "Beyond Benchmarks: DEA Study of Kansas Farm Productivity." *Southern Agricultural Economics Association's 2015 Annual Meeting*. Atlanta, Georgia, 2015. 20.
- R Core Team. "R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing, Vienna, Austria*. 2015. <http://www.R-project.org/>.
- Roth, Sarah, and Jeffrey Hyde. *Partial Budgeting for Agricultural Businesses*. Extension Publication, University Park, PA: Pennsylvania State University, 2002.
- Schimmelpfennig, David, and Robert Ebel. "On the Doorstep of the Information Age: Recent Adoption of Precision Agriculture." *EIB-80*. U.S. Department of Agriculture, Economic Research Service, August 2011.
- Simar, Leopold, and Paul W Wilson. "Estimation and Inference in Two-Stage, Semi-parametric Models of Production Processes." *Journal of Economics* 136 (2007): 31-64.
- Simar, Leopold, and Paul W Wilson. "Two-Stage DEA: Caveat Emptor." *Journal of Productivity Analysis*, 2011.
- Stacy, E W. "A Generalization of the Gamma Distribution." *The Annal of Mathematical Statistics* 33, no. 3 (September 1962): 1187-1192.
- The R Foundation. *About R: What is R?* March 22, 2015. <http://www.r-project.org/about.html> (accessed March 22, 2015).
- . *Home Page: The R Project for Statistical Computing*. March 22, 2015. www.r-project.org (accessed March 22, 2015).
- Trautmann, Heike, Detlef Steuer, Olaf Mersmann, and Bjorn Bornkamp. "truncnorm: Truncated Normal Distribution." *R package version 1.0-7*. 2014.
- US Department of Agriculture. "2012 Census of Agriculture County Profile for Clay County Kansas." 2012. <http://www.agcensus.usda.gov> (accessed February 17, 2015).
- Watkins, K Bradley, Tatjana Hristovska, Ralph Mazzanti, and Charles E Wilson, Jr. "Measuring Technical, Allocative, and Economic Efficiency of Rice Production in Arkansas using Data Envelopment Analysis." *Southern Agricultural Economics Association's 2013 Annual Meeting*. Orlando, 2013.
- Whelan, Brett, and James Taylor. *Precision Agriculture for Grain Production Systems*. Collingwood AUS: CSIRO Publishing, 2013.
- Whitacre, Brian E, Tyler B Mark, and W Terry Griffin. "How Connected are Our Farms?" *Choices*, 3rd Quarter 2014.

Xue, Mei, and Patrick T Harker. "Ranking DMUs with Infeasible Super-Efficiency DEA Models." *Managment Science* 48, no. 5 (May 2002): 705-710.

APPENDIX A: R SCRIPT FOR DATASET SIMULATION

A.1 Introduction

This appendix outlines the script utilized in the computer generation of the county used in the analysis. The necessity to simulate the data rather than use observed values came as a result of the immense survey which would be required to include actual observed data. As adoption of big data among organizations grows and customer data increases, this data will become more readily available to conduct this analysis using empirical values. When the required observations become available, they can be substituted for the script found in this appendix.

A.2 Outline of Headers Found in Script

The script used for generating farm data is broken into five parts: generation of farms (establishing acreage), breaking farm acreage into commodity groups, assigning precision technology adoption attributes, establishing yield for each commodity, and distributing a per-acre cost of inputs across the farms. The establishment of yield and distribution of per-acre cost of inputs will be designated on a normal distribution. The simulation of values should be substituted for real observed data whenever possible to generate the most realistic model for analysis.

The script notated in the next section will generate the same values every time unless “`set.seed(3414500)`” is removed. Also, the library package “`truncnorm`” (Trautmann, et al. 2014) should be installed in R in order for the generation of the farms to take place on a truncated normal distribution as scripted.

A.3 Dataset Simulation Script

```
set.seed(3414500)
library(truncnorm)
```

```

#2012 Ag Census Clay County, Kansas Attributes
c1<-rtruncnorm(n = 20, a = 1, b = 9, mean = 1, sd = 50)
c2<-rtruncnorm(n = 61, a = 10, b = 49, mean = 10, sd = 100)
c3<-rtruncnorm(n = 119, a = 50, b = 179, mean = 50, sd = 200)
c4<-rtruncnorm(n = 139, a = 180, b = 499, mean = 180, sd = 1000)
c5<-rtruncnorm(n = 71, a = 500, b = 999, mean = 570, sd = 1000)
c6<-rtruncnorm(n = 131, a = 1000, b = 10000, mean = 1000, sd = 1085)
acres<-c(c1, c2, c3, c4, c5, c6)
acres<-round(acres)

#Commodity Attributes
dat.sim<-data.frame(Total=acres)

#Corn
cn<-dat.sim$Corn<-ifelse(dat.sim$Total > 999,
abs(rnorm(541)*0.215*dat.sim$Total),0)

#Soybeans
sb<-dat.sim$Soybeans<-ifelse(dat.sim$Total > 999, dat.sim$Total*.43,
dat.sim$Total*0.333)

#Grain Sorghum
gs<-dat.sim$Sorghum<-ifelse(dat.sim$Total < 1000, dat.sim$Total*0.333,
abs(rnorm(541)*.033*dat.sim$Total))

#Wheat
dat.sim$Wheat=dat.sim$Total-cn-sb-gs
dat.sim$Wheat<-ifelse(dat.sim$Wheat<0,0,dat.sim$Wheat)

#Precision Attributes
#40% of total acres have a yield monitor
dat.sim$YIELDMONITOR<-ifelse(dat.sim$Total > 500, rbinom(541, 1, .41),0)
aggregate(Total~YIELDMONITOR, data=dat.sim, sum)
sum(dat.sim$Total)
aggregate(Total~YIELDMONITOR, data=dat.sim, sum)[2,2]/sum(dat.sim$Total)

#20% of total acres have Auto Guidance
dat.sim$AUTOGUIDANCE<-ifelse(dat.sim$Total > 500, rbinom(541, 1,
0.238),0)
aggregate(Total~AUTOGUIDANCE, data=dat.sim, sum)
sum(dat.sim$Total)
aggregate(Total~AUTOGUIDANCE, data=dat.sim, sum)[2,2]/sum(dat.sim$Total)

#10% of total acres have VR applications
dat.sim$VRT<-ifelse(dat.sim$Total > 100, rbinom(541, 1, 0.0904),0)
aggregate(Total~VRT, data=dat.sim, sum)
sum(dat.sim$Total)
aggregate(Total~VRT, data=dat.sim, sum)[2,2]/sum(dat.sim$Total)

#10% of total acres have a Yield Monitor, Auto Guidance, and use Big Data
dat.sim$BIGDATA<-ifelse(dat.sim$YIELDMONITOR==1 &
dat.sim$AUTOGUIDANCE==1, rbinom(541, 1, 0.923),0)
aggregate(Total~BIGDATA, data=dat.sim, sum)
sum(dat.sim$Total)
aggregate(Total~BIGDATA, data=dat.sim, sum)[2,2]/sum(dat.sim$Total)

```

```

#5% of acres are farmed with YM, Auto Guidance, & Big Data utilizing
Telematics
dat.sim$TELEMATICS<-ifelse(dat.sim$BIGDATA==1, rbinom(541, 1, .155),0)
aggregate(Total~TELEMATICS, data=dat.sim, sum)
sum(dat.sim$Total)
aggregate(Total~TELEMATICS, data=dat.sim, sum)[2,2]/sum(dat.sim$Total)

#Corn Yield
cn.sim<-rtruncnorm(n = 541, a = 5, b = 250, mean = 121.31, sd = 10)
cn.yld<-abs(cn.sim)
cn.vrt<-ifelse(dat.sim$VRT==1, cn.yld*rnorm(1, mean = 0.15, sd = 0.25),0)
cn.ag<-ifelse(dat.sim$AUTOGUIDANCE==1, cn.yld*rnorm(1, mean = 0.2, sd =
0.25),0)
cn.dat<-ifelse(dat.sim$BIGDATA==1, cn.yld*rnorm(1, mean = 0.15, sd =
0.25),0)
dat.sim$Corn.Yield<-abs(cn.yld+cn.vrt+cn.ag+cn.dat)
dat.sim$Corn.Yield

#Soybeans yield
sb.sim<-rtruncnorm(n = 541, a = 5, b = 85, mean = 37.84, sd = 30)
sb.yld<-abs(sb.sim)
sb.vrt<-ifelse(dat.sim$VRT==1, sb.yld*rnorm(1, mean = 0.15, sd = 0.25),0)
sb.ag<-ifelse(dat.sim$AUTOGUIDANCE==1, sb.yld*rnorm(1, mean = 0.2, sd =
0.25),0)
sb.dat<-ifelse(dat.sim$BIGDATA==1, sb.yld*rnorm(1, mean = 0.15, sd =
0.25),0)
dat.sim$Soybeans.Yield<-abs(sb.yld+sb.vrt+sb.ag+sb.dat)
dat.sim$Soybeans.Yield

#Sorghum Yield
gs.sim<-rtruncnorm(n = 541, a = 5, b = 150, mean = 88.1, sd = 50)
gs.yld<-abs(gs.sim)
gs.vrt<-ifelse(dat.sim$VRT==1, gs.yld*rnorm(1, mean = 0.15, sd = 0.25),0)
gs.ag<-ifelse(dat.sim$AUTOGUIDANCE==1,gs.yld*rnorm(1,mean=0.2,sd=
0.25),0)
gs.dat<-ifelse(dat.sim$BIGDATA==1, gs.yld*rnorm(1, mean = 0.15, sd =
0.25),0)
dat.sim$Sorghum.Yield<-abs(gs.yld+gs.vrt+gs.ag+gs.dat)
dat.sim$Sorghum.Yield

#Wheat Yield
wh.sim<- rtruncnorm(n = 541, a = 5, b = 120, mean = 40.85, sd = 20)
wh.yld<-abs(wh.sim)
wh.vrt<-ifelse(dat.sim$VRT==1, wh.yld*rnorm(1, mean = 0.15, sd = 0.25),0)
wh.ag<-ifelse(dat.sim$AUTOGUIDANCE==1, wh.yld*rnorm(1, mean = 0.2, sd =
0.25),0)
wh.dat<-ifelse(dat.sim$BIGDATA==1, wh.yld*rnorm(1, mean = 0.15, sd =
0.25),0)
dat.sim$Wheat.Yield<-abs(wh.yld+wh.vrt+wh.ag+wh.dat)
dat.sim$Wheat.Yield

#Gross Revenue
cn.rev<-dat.sim$Corn.Revenue<-ifelse(dat.sim$Corn > 1,
(dat.sim$Corn.Yield*dat.sim$Corn*3.60),0)
sb.rev<-dat.sim$Soybeans.Revenue<-ifelse(dat.sim$Soybeans > 0.1,
(dat.sim$Soybeans.Yield*dat.sim$Soybeans*8.73),0)

```

```

gs.rev<-dat.sim$Sorghum.Revenue<-ifelse(dat.sim$Sorghum > 0.1,
(dat.sim$Sorghum.Yield*dat.sim$Sorghum*3.95),0)
wh.rev<-dat.sim$Wheat.Revenue<-ifelse(dat.sim$Wheat > 0.1,
(dat.sim$Wheat.Yield*dat.sim$Wheat*5.02),0)

#Corn Inputs
cn.input<-ifelse(dat.sim$Corn > 1, (dat.sim$Corn*298.05),0)
ag.cninput<-ifelse(dat.sim$AUTOGUIDANCE==1, cn.input*rnorm(1, mean =
0.05, sd = 0.25),0)
dat.cninput<-ifelse(dat.sim$BIGDATA==1, cn.input*rnorm(1, mean = 0.15, sd
= 0.25),0)
tele.cninput<-ifelse(dat.sim$TELEMATICS==1, cn.input*rnorm(1, mean = 0.1,
sd = 0.25),0)
dat.sim$Corn.Input<-abs(cn.input-ag.cninput-dat.cninput-tele.cninput)
dat.sim$Corn.Input

#Soybeans Inputs
sb.input<-dat.sim$Soybeans.Input<-ifelse(dat.sim$Soybeans > 0.1,
(dat.sim$Soybeans*158.54),0)
ag.sbinput<-ifelse(dat.sim$AUTOGUIDANCE==1, sb.input*rnorm(1, mean =
0.05, sd = 0.25),0)
dat.sbinput<-ifelse(dat.sim$BIGDATA==1, sb.input*rnorm(1, mean = 0.15, sd
= 0.25),0)
tele.sbinput<-ifelse(dat.sim$TELEMATICS==1, sb.input*rnorm(1, mean = 0.1,
sd = 0.25),0)
dat.sim$Soybeans.Input<-abs(sb.input-ag.sbinput-dat.sbinput-tele.sbinput)
dat.sim$Soybeans.Input

#Sorghum Inputs
gs.input<-dat.sim$Sorghum.Input<-ifelse(dat.sim$Sorghum > 0.1,
(dat.sim$Sorghum*186.95),0)
ag.gsinput<-ifelse(dat.sim$AUTOGUIDANCE==1, gs.input*rnorm(1, mean =
0.05, sd = 0.25),0)
dat.gsinput<-ifelse(dat.sim$BIGDATA==1, gs.input*rnorm(1, mean = 0.15, sd
= 0.25),0)
tele.gsinput<-ifelse(dat.sim$TELEMATICS==1, gs.input*rnorm(1, mean = 0.1,
sd = 0.25),0)
dat.sim$Sorghum.Input<-abs(gs.input-ag.gsinput-dat.gsinput-tele.gsinput)
dat.sim$Sorghum.Input

#Wheat Inputs
wh.input<-dat.sim$Wheat.Input<-ifelse(dat.sim$Wheat > 0.1,
(dat.sim$Wheat*132.61),0)
ag.whinput<-ifelse(dat.sim$AUTOGUIDANCE==1, wh.input*rnorm(1, mean =
0.05, sd = 0.25),0)
dat.whinput<-ifelse(dat.sim$BIGDATA==1, wh.input*rnorm(1, mean = 0.15, sd
= 0.25),0)
tele.whinput<-ifelse(dat.sim$TELEMATICS==1, wh.input*rnorm(1, mean = 0.1,
sd = 0.25),0)
dat.sim$Wheat.Input<-abs(wh.input-ag.whinput-dat.whinput-tele.whinput)
dat.sim$Wheat.Input

#Net Revenue by Crop
cn.net<-dat.sim$Corn.NET<-cn.rev-dat.sim$Corn.Input
sb.net<-dat.sim$Soybeans.NET<-sb.rev-dat.sim$Soybeans.Input
gs.net<-dat.sim$Sorghum.NET<-gs.rev-dat.sim$Sorghum.Input
wh.net<-dat.sim$Wheat.NET<-wh.rev-dat.sim$Wheat.Input

```



```
#Total Farm Net Revenue
farm.net<-dat.sim$Total.Net<-cn.net+sb.net+gs.net+wh.net

#Total Farm Net Revenue/Acre
farm.netacre<-dat.sim$Total.NetAcre<-farm.net/dat.sim$Total
```

APPENDIX B: R SCRIPT FOR DATA ENVELOPMENT ANALYSIS

B.1 Introduction

This appendix outlines the script utilized in conducting a benchmarking data envelopment analysis. In order for this script to be applicable, it must be used in conjunction with a dataset using headers similar to the dataset generated in Appendix A. This analysis requires the installation of the “**Benchmarking**” (Bogetoft and Otto 2014) package for data envelopment and the “**censReg**” (Henningsen 2013) package for the Tobit Censored Regression. This script also includes the charting of the analysis for all four commodities as well as all five PA technology parameters.

B.2 Dataset Analysis Script

```
library(Benchmarking)
dat<-dat.sim
X<-dat$Total
Y<-dat$Total.Net

dea.vrio<-dea(X,Y, ORIENTATION="in", RTS="vrs")
dea.vrio

dea.crio<-dea(X,Y, ORIENTATION="in", RTS="crs")
dea.crio

dea.se<-eff(dea.crio)/eff(dea.vrio)
dat.sim$se<-dea.se

eff.metric<-dat.sim[,29]
library(truncreg)

library(censReg)
DEAcensReg<-
censReg(eff.metric~YIELDMONITOR+AUTOGUIDANCE+VRT+BIGDATA+TELEMATICS,
right=1)
summary(DEAcensReg)

dat.sim.BD<-subset(dat.sim, BIGDATA==1)
dat.sim.AG<-subset(dat.sim, AUTOGUIDANCE==1)
dat.sim.TEL<-subset(dat.sim, TELEMATICS==1)
dat.sim.VRT<-subset(dat.sim, VRT==1)
dat.sim.YM<-subset(dat.sim, YIELDMONITOR==1)
dat.sim.CN<-subset(dat.sim, cn.net>0)
dat.sim.SB<-subset(dat.sim, sb.net>0)
dat.sim.GS<-subset(dat.sim, gs.net>0)
dat.sim.WH<-subset(dat.sim, wh.net>0)
```

```

png("Total_Data.png", width = 1600, height = 1200, res = 300,
family="serif")
dea.plot(x=dat.sim$Total, y=dat.sim$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresBD.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.BD$Total, y=dat.sim.BD$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresAG.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.AG$Total, y=dat.sim.AG$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresTEL.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.TEL$Total, y=dat.sim.TEL$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresVRT.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.VRT$Total, y=dat.sim.VRT$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresYM.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.YM$Total, y=dat.sim.YM$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresCN.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.CN$Total, y=dat.sim.CN$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresSB.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.SB$Total, y=dat.sim.SB$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresGS.png", width = 1600, height = 1200, res =
300, family="serif")
dea.plot(x=dat.sim.GS$Total, y=dat.sim.GS$Total.NetAcres, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

png("DEA_NetAcres_TotalAcresWH.png", width = 1600, height = 1200, res =
300, family="serif")

```

```

dea.plot(x=dat.sim.WH$Total, y=dat.sim.WH$Total.NetAcre, RTS="vrs",
ORIENTATION="in-out", xlab="Acreage", ylab="Output")
dev.off()

dea.vrio.BD<-dea(X=dat.sim.BD$Total, Y=dat.sim.BD$Total.Net,
ORIENTATION="in", RTS="vrs")

dea.crio.BD<-dea(X=dat.sim.BD$Total, Y=dat.sim.BD$Total.Net,
ORIENTATION="in", RTS="crs")

BD.se<-eff(dea.crio.BD)/eff(dea.vrio.BD)

summary(BD.se)

dea.vrio.YM<-dea(X=dat.sim.YM$Total, Y=dat.sim.YM$Total.Net,
ORIENTATION="in", RTS="vrs")

dea.crio.YM<-dea(X=dat.sim.YM$Total, Y=dat.sim.YM$Total.Net,
ORIENTATION="in", RTS="crs")

YM.se<-eff(dea.crio.YM)/eff(dea.vrio.YM)

summary(dea.vrio.YM)

dea.vrio.TM<-dea(X=dat.sim.TEL$Total, Y=dat.sim.TEL$Total.Net,
ORIENTATION="in", RTS="vrs")

dea.crio.TM<-dea(X=dat.sim.TEL$Total, Y=dat.sim.TEL$Total.Net,
ORIENTATION="in", RTS="crs")

TM.se<-eff(dea.crio.TM)/eff(dea.vrio.TM)

dea.vrio.VR<-dea(X=dat.sim.VRT$Total, Y=dat.sim.VRT$Total.Net,
ORIENTATION="in", RTS="vrs")

dea.crio.VR<-dea(X=dat.sim.VRT$Total, Y=dat.sim.VRT$Total.Net,
ORIENTATION="in", RTS="crs")

VR.se<-eff(dea.crio.VR)/eff(dea.vrio.VR)

dea.vrio.AG<-dea(X=dat.sim.AG$Total, Y=dat.sim.AG$Total.Net,
ORIENTATION="in", RTS="vrs")

dea.crio.AG<-sdea(X=dat.sim.AG$Total, Y=dat.sim.AG$Total.Net,
ORIENTATION="in", RTS="crs")

AG.se<-eff(dea.crio.AG)/eff(dea.vrio.AG)

```

```

summary4thesis<-matrix(0, nrow=6, ncol=4)

colnames(summary4thesis)<-c("Mean", "Std. Dev.", "Min", "Max")

rownames(summary4thesis)<-c("Scale Efficiency", "Big Data SE", "Yield
Monitor SE", "Telematics SE", "VRT SE", "Auto Guidance SE" )

summary4thesis[1,1]<-mean(dea.se)

summary4thesis[1,2]<-sd(dea.se)

summary4thesis[1,3]<-min(dea.se)

summary4thesis[1,4]<-max(dea.se)

summary4thesis[4,1]<-mean(TM.se)

summary4thesis[4,2]<-sd(TM.se)

summary4thesis[4,3]<-min(TM.se)

summary4thesis[4,4]<-max(TM.se)

summary4thesis[2,1]<-mean(BD.se)

summary4thesis[2,2]<-sd(BD.se)

summary4thesis[2,3]<-min(BD.se)

summary4thesis[2,4]<-max(BD.se)

summary4thesis[3,1]<-mean(YM.se)

summary4thesis[3,2]<-sd(YM.se)

summary4thesis[3,3]<-min(YM.se)

summary4thesis[3,4]<-max(YM.se)

summary4thesis[5,1]<-mean(VR.se)

summary4thesis[5,2]<-sd(VR.se)

summary4thesis[5,3]<-min(VR.se)

summary4thesis[5,4]<-max(VR.se)

summary4thesis[6,1]<-mean(AG.se)

summary4thesis[6,2]<-sd(AG.se)

summary4thesis[6,3]<-min(AG.se)

summary4thesis[6,4]<-max(AG.se)

round(summary4thesis, 4)

```