SOYBEAN QUALITY AND MARKET FACTOR INFLUENCE ON NORTH DAKOTA ORIGIN BASIS VALUES

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Ryan Wayne Thompson

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Agribusiness and Applied Economics

January 2020

Fargo, North Dakota

North Dakota State University
Graduate School

Title

SOYBEAN QUALITY AND MARKET FACTOR INFLUENCE ON

NORTH DAKOTA ORIGIN BASIS VALUES

By

Ryan Wayne Thompson

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. David Bullock
Chair

Dr. William Wilson

Dr. Fariz Huseynov

Approved:

04/07/2020
Date

Dr. William Nganje
Department Chair

ABSTRACT

In North Dakota, origin soybean basis values place risk on market participants involved with trading the commodity. Researchers study market factors that influence origin basis values in order to better understand basis movements. Quality is an important component of basis values, but soybean quality in the form of crude protein and essential amino acid content may be implicitly reflected in the basis rather than explicitly in the flat price. A panel data set is created comprised of ten years of data ranging from 2009-2018 and across eight North Dakota agricultural statistics districts. This thesis utilizes principal component analysis to identify principal components among essential amino acid variables. Panel regression models using fixed effects are employed to identify the influence soybean quality and other market factors have on origin soybean basis values in North Dakota. Results indicate that essential amino acids and protein implicitly influence North Dakota origin soybean basis values.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. David Bullock for his direction and help throughout the entire research process. The knowledge and expertise he provided is greatly appreciated. I would like to thank Dr. William Wilson for his continued guidance throughout my time at North Dakota State University. He is the one who sparked my graduate school and career interests and helped me pursue them. I would also like to thank Dr. Fariz Huseynov for serving on my thesis committee and providing his expertise. Thank you to Dr. Prithviraj Lakkakula for providing his expertise on the topics in this thesis.

I would like to thank the members of the Agribusiness and Applied Economics department at North Dakota State University. The exceptional atmosphere provided by the faculty and staff made learning, research, and personal development second nature. Having this type of environment is what makes North Dakota State University a premier agricultural institution.

Thank you to the North Dakota Soybean Council who provided the necessary funding for this research. Their commitment to furthering the literature surrounding soybeans in North Dakota positively impacts the lives of countless people throughout the industry.

Finally, I would like to thank my family. They have given me a strong foundation from which I have been able to pursue my goals. They continue to lead by example and have provided me with valuable life lessons. They have continually pushed me to reach my full potential both in academics and in life.

TABLE OF CONTENTS

ABSTRACT……………………………………………………………………………….. iii

ACKNOWLDEGEMENTS………………………………………………………………... iv

LIST OF TABLES…………………………………………………………..…………….. vii

LIST OF FIGURES…………………………………………………………..……..…. viii

CHAPTER 1. INTRODUCTION…………………………………………………...… 1

1.1. Overview………....……………………………………………………………. 1

1.2. Problem and Objectives……………...………………….…………..…...…………….. 1

1.3. Methods and Procedures………………...……..…………………………….… 3

1.4. Organization…………………..………………………………………………… 4

CHAPTER 2. REVIEW OF LITERATURE……………………………………..…………. 5

2.1. Introduction……..…………………………..……………………………….……. 5

2.2. Basis Forecasting……..…………………...………………..…………………….…. 5

2.3. Market Factors……………………..………..………………………………....……. 7

2.4. Soybean Quality…………………..……………………..………………...…..…….. 9

2.5. Summary…………………...………...……….………………………...…………. 11

CHAPTER 3. THEORETICAL FRAMEWORK…..……………………………………12

3.1. Introduction……………….………………………………………………………... 12

3.2. Basis Theory…………….....…………………………………………..……….…. 13

3.3. Soybean Quality……………………..…...………………...…………………….……… 17

3.4. Summary…………………………………………………………………..………….. 18

CHAPTER 4. EMPIRICAL METHODS……………………………..………..……………. 20

4.1. Introduction…………..………………………………………………….………. 20

4.2. Panel Regression Data……….……….………………………………….………… 21

4.3. Data Adjustments………......……………...…………………………………..… 23

4.4. Empirical Model 24

4.5. Principal Component Analysis.. 25

4.6. Correlation Matrices.. 26

4.7. Model Specification.. 26

4.8. Summary..27

CHAPTER 5. RESULTS AND ANALYSIS.. 29

5.1. Introduction.. 29

5.2. Overview of Results.. 29

5.3. Data Adjustments.. 30

5.4. PCA Results.. 31

5.5. Correlation Matrix Results.. 33

5.6. Model Specification Results.. 34

5.7. Panel Regression Results... 36

5.8. Summary... 39

CHAPTER 6. CONCLUSION.. 41

6.1. Introduction... 41

6.2. Problem and Objectives... 41

6.3. Methodology and Model.. 42

6.4. Overview of Results.. 43

6.5. Contribution to Literature... 45

6.6. Implications... 46

6.7. Continued Research... 46

REFERENCES.. 48

LIST OF TABLES

Table Page

4.1. Agricultural Statistic District Origins…………………………………………………… 23

4.2. Model Variables………………………………………………………………………. 25

5.1. Essential Amino Acid PCA Squared Cosines…………………………………………… 31

5.2. PCA Factor Loadings…………………………………………………………………..…33

5.3. Pearson Correlation Coefficients…………...……………………………………………… 34

5.4. Pearson Correlation P-Values……………...…………………………………………… 34

5.5. Model Specification P-Values……….....…...……………………………………………. 35

5.6. Panel Regression Results………………...………………………………..……………… 36

5.7. Empirical Results………………………..……...…………………………..…………..39

LIST OF FIGURES

Figure Page

4.1. North Dakota Agricultural Statistic Districts...……………………………………..……..… 22

5.1. Ten Year Essential Amino Acid Factor Scores…………………...………………...……... 30

5.2. Nine Year Essential Amino Acid Factor Scores……...……………………………………. 31

5.3. Correlation Circle……………………………...……………………..……………………... 33

CHAPTER 1: INTRODUCTION

1.1. Overview

Soybeans are vital to agriculture in North Dakota and the Upper Midwest region of the United States. According to the United States Department of Agriculture (USDA), North Dakota alone produced over 239 million bushels in the 2018 crop year. Among other factors, growing conditions and soil types are suitable for the commodity in this region. The continued production is in part due to the foreign and domestic demand for the commodity with efficient methods of shipping to export markets.

1.2. Problem and Objectives

Soybeans are produced across a large portion of the Midwest and carry spatial differences in terms of quality. For years, market participants and the marketing system have used crude protein content as a measure of soybean quality. Foreign material, heat damage and other quality aspects are other common quality characteristics used by market participants to assess premiums or discounts. Though these other characteristics are important, crude protein and essential amino acid (EAA) content are the two quality variables analyzed in this study. Crude protein is used as a proxy of value; however, research has shown there are more accurate measures of soybean quality. According to Hertsgaard (2015) essential amino acid content is a better indicator of soybean quality than crude protein content. He found that soybeans with lower levels of crude protein content have higher levels of the five most critical amino acids lysine, cystine, threonine, methionine, and tryptophan. Soybeans in the Upper Midwest had lower crude protein content, but higher levels of these EAA. These findings are beneficial to soybean market participants in the Upper Midwest due to the desirability of EAA in soybean. Soybeans are processed into soybean meal which is a widely used feedstuff. Soybean meal is important for the nutritional benefit it contributes to feed rations fed to livestock, poultry, and other animals. Soybean meal provides the animal with the large majority of the EAAs in their diet at a reasonable cost. Essential amino acids are amino acids that cannot be synthesized by animal`s body. For this reason, nutritionists must include these EAAs in the animal's diet in order to grow and lead a healthy life.

Basis is a crucial market mechanism that help coordinate the movement along the supply chain from producer to end user. Basis in its simplest form is defined as the cash price of a commodity minus its futures price. Basis volatility places substantial risk on both buyers and sellers of the commodity. With volatility presenting basis risk there has been many studies by researchers attempting to forecast basis movements. Most basis forecasting models use moving averages or other time series methods to forecast basis values. Though forecasting models are common, having the knowledge of what factors influence basis is important.

There are numerous variables that have been found to influence soybean basis volatility. The cost of shipping a commodity from origin to destination is a key variable that is factored into origin basis values. When shipping costs for elevators increase the elevator may need to adjust basis levels to cover the added cost. The export basis at the Pacific Northwest (PNW) represents supply and demand for the commodity at the export terminal. North Dakota origins rely on the PNW as a destination from which soybeans can be exported. Futures spreads are used as a variable in this study as they are an indication of market supply and demand at origin facilities.

Identifying relationship North Dakota origin basis values have with soybean quality and other market factors are objectives of this thesis. The first objective of this thesis is to identify the influence soybean protein and essential amino acid content has on origin basis values in North Dakota. While crude protein content has been used as a measure of soybean quality, literature has shown that essential amino acid content in soybeans to be a better quality indicator. With research showing North Dakota to have higher levels of EAA, do these quality characteristics influence origin basis in the state? The second objective is to identify the influence certain market factors have on origin basis values in North Dakota. Understanding these relationships allows market participants to better understand and mitigate their basis risk.

1.3. Methods and Procedures

This research uses a wide range of variables that form a panel data set. A panel dataset is comprised of time series and cross-sectional data. The data in this study ranges from 2009-2018 providing ten years of historical data. The cross-sectional units are across eight of the nine agricultural statistic districts that North Dakota is divided into. Daily origin basis values from one origin in each of the eight statistical districts are gathered. One market year EAA value is gathered in each statistical district for the five EAAs analyzed in this study. One market year crude protein content value is also gathered in each of the statistical districts. The remaining data: secondary car values, Pacific Northwest export basis values, and futures spreads are constant across the cross sections. Necessary data adjustments are made in order to obtain a complete dataset.

A model is developed in order to identify the significance of the market factor and quality variables on origin basis values in North Dakota. The model is first applied to market year data in which variables are averaged across the soybean market that ranges from September 1st through August 31st. The model is also applied to quarterly data in which the data is averaged by corresponding quarter with the first quarter starting on September 1st.

A common variable reduction technique called principal components analysis (PCA) is applied to the amino acid variables in this study to identify common underlying factors among the variables known as principal components. PCA attempts to limit the number of correlated variables while being able to retain the same amount of information in the form of said principal components.

Model specification is needed when researching variables in a panel data set. Two model specification tests are used for determining whether a fixed or random effects model should be used for the model. The Hausman specification and Breusch-Pagan tests are the two tests used in this study to do so. The results of the model specification tests are interpreted through p-values and determine that the fixed effects model is appropriate for estimation.

1.4. Organization

This thesis is organized into six chapters. Chapter 1 introduces the topic, defines the problem, and briefly describes the research. Chapter 2 examines the literature of previous studies in association with this study. Chapter 3 is theory and conceptual framework chapter. This chapter describes the theory of basis and its relationship with the variables in this study. This chapter also describes the importance of the quality characteristics in soybeans. Chapter 4 is the empirical models. This chapter describes the models and the methods used to conduct the research. The dataset, its collection, and organization are discussed in this section. Analytical methods and other preliminary tests are explained. Chapter 5 describes the results of the preliminary testing as well as the results of the model defined in this study. Chapter 6 summarizes the study. It defines how the research addresses the problems, discusses significant results, identifies implications of the research, and provides avenues for continued research.

CHAPTER 2. REVIEW OF LITERATURE

2.1. Introduction

This chapter presents the review of relevant literature. The first section of the literature review focuses on soybean basis modeling and forecasting research. There are several basis forecasting studies that relate to this thesis. It is important to understand the methods of basis forecasting as market participants are constantly trying to predict basis movements. The second section of the literature review focuses on explanatory factors that have been found to influence soybean basis. For the purpose of this study, it is important to identify and understand the variables that affect soybean basis through previous research.

2.2. Basis Forecasting

Forecasting commodity basis using a specific model or method is a common practice by commodity market participants. Basis values are important to forecast for producers, elevators, and end users looking to mitigate basis risk. According to Sanders and Manfredo (2006), understanding basis behavior is important for placing effective hedges. Small changes in basis values can lead to large financial gains or losses for market participants handling large volumes of a commodity.

Most basis forecasting models employ time series methods to predict future values. One reason for this is that collecting data for these models is relatively easy as there is usually a large amount of historical time series basis data for elevators or export terminal locations. With basis forecasting being so important, there is much speculation as to which model is the most accurate. Multiple studies have been conducted to identify model accuracy as well as the correct length of time series data to use.

Using moving averages is a very common and simple method of forecasting values in a time series data set. Moving average techniques are used for forecasting basis values by market participants in order to get a better idea of future basis movements. There have been numerous studies analyzing which type of moving average provides the most accurate forecast. Taylor et al. (2002) looked at determining the optimal number of years to be used when forecasting basis values of different commodities using moving averages. Having a moving average model that utilizes the optimal number of years would improve the success rate for market participants forecasting basis.

Taylor et al. (2002) also wanted to determine if they could improve their forecasting accuracy by incorporating current market data. Seven different historical averages were used along with a naïve forecast and a historical average including current basis information. The current basis information is considered as the deviation by the current basis values from its historical average. These models were applied using soybean, corn, and milo basis at six elevators in Kansas. Mean absolute errors were used to standardize and interpret their results. They found that using the previous year or the one-year average to forecast basis worked better than the longer-term averages. The one exception to this was for wheat in which a five-year average performed better. Using shorter moving averages when forecasting soybean basis is conceptually practical due to the volatility of soybean markets. They also found that using current basis information increases the accuracy of the forecasts.

Sanders and Manfredo (2006) took a different approach to forecasting basis and applied their study to origin soybean basis**.** Rather than looking at one model and adjusting the duration of the time series data like Taylor et al., their study looks at six different types of time series models. Sanders and Manfredo found that of the six time-series models the auto regressive moving average (ARMA) model was most accurate for soybeans followed by the year ago (YAG) in which the forecasted basis at a particular time would be the same as the basis from exactly one year prior.

2.3. Market Factors

Forecasting models using past basis values fail to consider changing external variables that impact basis. Rather than forecasting, other research has been done to identify market factors that influence basis values. There are several studies that identify these market factors and determine their significance on origin soybean basis. Examples of these variables are futures spreads, shipping costs, and export basis values.

Soybean buyers such as elevators and other grain handling facilities look at these and other explanatory variables before setting the basis at their location. Soybean producers must also look at these factors when it comes to marketing their soybean crop. For example, low soybean yield for a region due to a drought may cause soybean basis to strengthen in that area because elevators need to fill their demand with a lower than expected supply. This is one example of how these explanatory variables are significant when determining origin soybean basis movements and why studying them is important.

Wilson and Dahl (2011) examined the relationships between market factors such as shipping costs and basis values. They found several explanatory variables to be significant in causing variability in origin soybean basis. A few of these variables include shipping costs, outstanding export sales, and the ratio of grain supplies to storage capacity. Increases in technology have caused the relationships between these variables and basis values to change over time. Logistical improvements such as the creation of the shuttle train system, covered hopper cars, and demurrage are examples of increased technology that have decreased the cost of rail shipping and secondary car market values.

Wilson and Dahl (2011) also go on to highlight important findings of their research. They emphasize that explanatory variables have different levels of significance depending on what region the origin is in. For example, origin soybean basis values in North Dakota will be more responsive to rail shipping costs than barge costs because there aren't any barge loading facilities, not even on the Red River. Conversely, the Minneapolis St. Paul area soybean basis values are more responsive to barge cost due to the large number of barge-loading facilities in that region.

Shipping cost is one of the major variables known to impact both export terminal and origin basis values. Though other variables are considered, grain merchandisers largely set their basis values to reflect these costs. Hart and Olson (2017) examined how shipping costs and transportation disruptions affect basis values. The results of their study show "increases in the tariff rates, fuel surcharges, and the costs of transportation in the secondary markets have a direct, inverse impact on the basis" (p.30). Variables such as shuttle train and ocean shipping costs were also highly significant having an inverse effect on local basis values. Increasing the logistical costs of transporting and handling grain typically causes grain merchandisers to weaken their basis in order to cover these costs.

Bullock and Wilson (2019) studied how 27 different explanatory variables impact soybean basis at the U.S. Gulf (USG) and Pacific Northwest (PNW) export terminals**.** Futures spreads, export competition, and logistical costs were examined on export terminal basis. They found that the Brazilian export terminal basis has the greatest impact on the Gulf and PNW average export basis values. When Brazilian export terminals change their soybean basis the

U.S. Gulf and PNW follow in order to remain competitive in the world market. Though their research looks at the significance of these explanatory variables on the export basis, certain variables in their study may have significance on origin soybean basis in North Dakota. Upper Midwest and North Dakota origins are dependent on Pacific Northwest export terminals as a destination market for their commodities. This suggests that basis changes at the Pacific Northwest export terminals could have significant impacts on the basis at North Dakota origins.

For the purposes of this study Chinese demand for United States soybeans is assumed to be reflected in the export basis levels at the Pacific Northwest. This assumption is made because Bullock and Wilson found that, “U.S. marketing year average soybean export basis levels are impacted by the level of soybean imports by China.” (p.27) Therefore, using United States soybeans imported by China as an explanatory variable is unnecessary.

2.4. Soybean Quality

Soybean protein content is an important quality indicator of soybeans grown in the United States and is important to consider when marketing soybeans both domestically and internationally. Soybeans that are exported or crushed domestically must meet quality requirements held by end users of soybeans. Soybean buyers prefer higher levels of protein because high protein soybeans are especially desirable for soybean crushing and feeding livestock. Though high protein soybeans are typically desirable, there are protein content level differentials spatially.

Thakur and Hurburgh (2007) studied quality differentials between United States soybean meal and soybean meal from other origins. Origins in their study included Brazil, Argentina, and other countries that produce significant amounts of soybeans each year. Their study analyzed essential amino acid content, protein content, and other quality variables associated with soybean meal. Thakur and Hurburgh (2007) examined the five most critical essential amino acids to an animal's diet. The lysine, methionine, threonine, cysteine and tryptophan content of the soybean samples from each origin are measured. They found United States soybeans to have superior levels of total essential amino acid content and inferior levels of protein compared to the other origins.

In his 2015 thesis, Hertsgaard studied protein and EAA content of soybeans in the Midwest. He found significant spatial variability in soybean quality characteristics. He stated that Upper Midwest soybeans have had lower crude protein content compared soybeans grown elsewhere in the United States. Though this may seem to be a problem, his study has found that lower crude protein levels may mean higher levels of essential amino acids which are desirable when looking for optimal feed rations. These results align with that of Thakur and Hurburgh (2007).

Crude protein content in soybeans have been used as a quality proxy of soybean value for several years, however research has found that there is a more precise measure of quality. When Hertsgaard et. al (2018) examine soybean quality, they concluded looking strictly at the crude protein values may be misleading. They found that rather than examining crude protein values, amino acids should be focused on to ensure the best value of soybeans and soybean meal is determined. Furthermore, Hersgaard et. al (2018) state, "Sophisticated livestock feed mixes are developed based on amino acids that make up proteins" (p. 267). Nutritionists target higher amino acid values rather than crude protein when looking to feed their livestock. "Five amino acids are essential to feeding formulations including cysteine, lysine, methionine, threonine, and tryptophan, and these are known as the essential amino acids (EAAs)." (Hertsgaard et. Al, 2018,

p. 267). This suggests that soybeans with low protein content which were previously thought to be less valuable may be more desirable due to their high EAA content.

2.5. Summary

Movements in the basis of a commodity at the origin are important to understand for market participants looking to minimize their basis risk. Review of relevant literature shows basis forecasting research that has been done for market participants to better predict basis movements. The literature also shows there are multiple explanatory variables that influence soybean basis. Transportation costs and basis at export terminals among other variables have been found to influence origin soybean basis. Knowing what factors cause the greatest amount of basis volatility would give market participants more information about origin basis in their respective areas of North Dakota.

Soybean quality characteristics such as EAA and protein have been studied due to their importance to end users of the commodity. Protein content has been an important industry indicator of quality; however, essential amino acids have been found to be even more valuable when creating optimal feed rations utilizing protein-rich soybeans. With EAA being the desirable nutritional product of soybeans, there needs to be an accurate understanding of how soybean protein and essential amino acid values impact soybean basis.

CHAPTER 3. THEORETICAL FRAMEWORK

3.1. Introduction

This chapter presents the theoretical framework that is the foundation for this model. The first section focuses on basis theory and the functional importance of basis to market participants. The second section focuses on the quality component of basis and the importance of essential amino acids in soybeans used in livestock feeding rations.

A large value is placed on the importance of market participants to be able to understand origin commodity basis and its movements. Hedgers looking to minimize risk need to understand basis movements so that they can better decide when to lift the hedge. Forecasting models such as moving averages are commonly used to predict basis movements by market participants in order to limit origin basis risk. Though these models are widely used, they only analyze past basis data making the forecast susceptible to changes in explanatory market factors. Changes in logistical costs and other market factors can influence basis values exposing market participants to higher amounts of risk. Wilson and Dahl (2011) emphasize that these explanatory variables have varying influences on basis depending on what region the origin is in. This leaves market participants trying to understand what explanatory variables have the largest influence on origin soybean basis as they try to mitigate risk caused by the variables.

Soybean protein content is a quality characteristic that is used as a proxy in the commodity merchandising industry. Hertsgaard (2015) determined that when it comes to soybean quality, EAA values rather than crude protein content are a more accurate predictor of soybean value. Nutritionists that mix livestock feeding rations target soybeans with high levels of essential amino acids as they are needed in the diet for optimal animal health and growth.

3.2. Basis Theory

Origin basis values serve a large purpose both in commodity marketing and as a market facilitator of grain. Commodity basis is generally comprised of four main components: time, location, form and random. The time component accounts for carrying the commodity into the future. The location component represents the shipping costs associated with moving the commodity between locations. The form component accounts for the quality of the commodity. Finally, the random component encompasses the remaining market factors such as supply and demand, etc. "One of the primary roles of the basis for a storable commodity is to coordinate the flow of the commodity from its location of production to its highest source of demand from a temporal, geographic, and form perspective" (Bullock and Wilson, 2019, p.1). The flow these commodities take from producer to origin and on to destination (or end user) are coordinated with the basis as an instrument for pricing.

Basis functions as a communication mechanism between commodity market participants. Basis levels influence the movement of commodities by signaling for market participants to move the commodity into storage or out into the market. For example, if an elevator needs to fill a shuttle train in the coming weeks and has low levels of inventory the elevator may increase the basis of the commodity in order to encourage delivery by producers to that location. Producers in the area experiencing low yield due to tough growing conditions is another reason elevators will strengthen their basis in order to encourage delivery to their location. Basis values can also be used by elevators to discourage the delivery of grain due to lack of storage capacity or other factors affecting a facility at any given time. "Simply, if there is a shortage of storage capacity, basis levels are lower, no doubt reflecting the impact of the shortage of storage capacity on local basis values" (Wilson and Dahl, 2011, p. 431). This emphasizes that changing supply and demand factors influence origin basis.

In normal market years origin basis values have seasonal characteristics that relate to the timing of certain reoccurring market factors and crop cycles. Basis values are usually weaker on average during harvest due to the large supply of grain being available to the market in a relatively small time period. These seasonal behaviors vary from year to year based on changing crop conditions, supply, and demand factors.

Logistical factors such as the costs of moving a commodity from origin to destination are also reflected in basis values. Elevators that utilize shuttle trains to ship their commodity have logistical costs such as rail tariffs, primary car values, and secondary car values. Rail tariffs are what shippers must pay to the railroad in order to use their line and are charged to shippers in dollars/car. The primary market is where railroads auction forward shuttle shipments. "The primary market, although with some variation across carriers, is the initial allocation of trains where shippers bid for rights to utilize a specified number of trains for a certain time period forward" (Lakkakula and Wilson, 2019, p.7). Secondary car markets allow companies to buy and sell shuttle contracts for a specific railroad. This market is generally more volatile due to the large number of companies trading contracts on one rail line. The value to shippers derived from these markets are called daily car values and are either a premium or discount to the rail tariffs that shippers must pay. The rail tariff plus the daily car (secondary car) values are the shipping cost incurred by the shipper. This does not include other logistical costs such as demurrage.

Elevators need to change basis values in accordance with changing shipping costs. An increase in shipping costs from an origin to a destination may cause origin basis to weaken for producers looking to deliver grain. This weakening of the basis allows for elevators to cover the increased transportation costs and is why elevators located further away from export terminals may have weaker basis values. Lakkakula and Wilson (2019) conducted a study on the effects of shipping costs on origin and destination basis values in Midwestern states between 2004 and 2016. They found that for every dollar increase in shipping cost there is a 19-cent decrease in origin basis for the origins in their study. Though only a handful of origins were studied, the research shows how significant shipping cost can be on origin basis. Country elevators (elevators not located on a major railway) that need to haul their grain to elevators located along one of the major railways have an extra cost that must be accounted for when determining their basis. These elevators typically ship using semi and trailer or use short-line railroads to move their commodity to an elevator that can load shuttle trains. The added cost of moving the commodity from country elevator to shuttle loader is usually reflected in the basis of the country elevator.

Basis in its simplest form is defined as the spot (cash) price of a commodity minus the futures price in a specific month. This is be written as:

$$B_o = C - F \tag{3.1}$$

Where B_o is the basis at an origin, C is the cash price at the same origin, and F is the futures price of the commodity in a specific futures month. Though this model calculates basis by showing the relationship between basis, cash price, and futures price of a specific commodity the model doesn't include explanatory variables and how they have been found to influence origin basis. Wilson et al. (2019) defines a simple origin basis model as:

$$B_o = (B_1 - (RR + 2^{nd})_1 - H_1) \tag{3.2}$$

Wilson et al (2019) define the formula, "where B_o is the origin basis, B_1 is the basis at terminal market 1, RR is the rail tariff to terminal market 1, and 2^{nd} is the value of freight in the secondary market and H_1 is the handling margin" (p.21) This basis model incorporates explanatory variables that influence origin basis values at an origin grain handling facility. In their model, shipping costs are represented by adding the rail tariff to the terminal market to the secondary rail market. Handling margin and shipping costs are subtracted from the terminal basis in order to derive the origin basis. Along with adjusting for shipping and handling costs of a commodity, basis values may also be adjusted so that elevators can remain competitive in their respective area. They go on to define an origin basis model that includes spatial competition between regional elevators. This model is defined as:

$$B_o = MAX\ [(B_1-(RR+2^{nd})_1-H_1),\ [(B_2-(RR+2^{nd})_2-H_2),\ldots,[(B_n-(RR+2^{nd})_n-H_n)] \qquad (3.3)$$

Wilson et al (2019) define the formula, “where values are as previously defined, and subscripts 2, …n are to represent values at competing terminal or export markets” (p.21). The model derives the best origin basis value offered in an area with “n” number of elevators. These models show import market factors influencing origin basis values. They also show different ways that basis can be derived depending on the export basis, shipping costs and handling margins affecting an origin.

Futures spreads have been found to influence commodity basis. Wider spreads in a normal market between futures contracts indicate that there is a large amount of a commodity entering the market which weakens the basis of the commodity due to supply and demand factors. Narrower future spreads indicate nearby demand which in turn will strengthen origin basis. Futures spreads communicate to market participants by letting them know if grain should be stored or sold into the market. If the cost to store grain is larger than the spread, then the market is looking for grain and it should be sold. If the spread is over a specific time period is larger than the storage cost over the same time period, then it is wise for market participants to store their commodity and capture the returns from storage. This is supported by Wilson and

Dahl (2011) when their research found that futures spreads significantly influence origin basis. In their research of 36 origins from 2004-2009, wider futures spreads lead to weaker basis levels at the origins.

3.3. Soybean Quality

Soybeans and soybean meal are desirable for their nutritional benefits when feeding livestock, poultry, and other animals. According to the Georgia State University Department of Poultry Science soybean meal provides these animals with around 80% of the amino acids in their diet. Soybean meal has many highly digestible essential amino acids including lysine, methionine, threonine and has high levels of cysteine. Compared to other oilseed meals there is more consistency in the digestibility of soybean meal.

Quality characteristics of soybeans such as essential amino acid content are important in the livestock, poultry, and swine industries. Certain EAAs are desired in soybeans more than others. Lysine, methionine, threonine, tryptophan, and cystine are critical amino acids that are important for animal health. EAAs are amino acids that cannot be synthesized naturally by livestock and other mammals. With insufficient levels of amino acids in the animal`s body there needs to be dietary amino acids within the animal`s feed ration. Without EAAs in an animal`s diet they cannot grow or develop to their full potential.

According to the Georgia State University Department of Poultry Science there are many factors that influence amino acid content in soybeans such as processing and storing the soybeans certain ways. Growing conditions and plant variety can also impact the amino acid content in soybeans. With growing conditions playing a part in the level of amino acids in soybeans there becomes spatial variability in amino acid content. Hertsgaard (2015) found crude protein values to be substantially lower in the northern Midwest states compared to the rest of the region. Though crude protein values were lower in the northern Midwest essential amino acid values were significantly higher. Hertsgaard used the Critical Amino Acid Value to show the spatial differences among amino acid content in soybeans. The CAAV consists of lysine, cysteine, methionine, threonine, and tryptophan which are the most important amino acids to livestock, poultry, and other animals. The CAAV is the summation of these amino acids as a percentage of the eighteen total amino acids. Hertsgaard`s research shows an inverse relationship between crude protein and amino acid content in soybeans meaning low crude protein content soybeans have high EAA content. North Dakota and the northern Midwest soybean producing states have been known to produce soybeans with low crude protein content which makes them viewed as less valuable. There are two pieces of information that show soybeans produced in North Dakota to be valuable based on quality: (a) research showing soybeans produced in North Dakota have high CAAV (EAAs) and (b) research showing essential amino acids are critical to feed-animal health and growth.

3.4. Summary

Basis movements can be caused by multiple market factors. These relationships were defined in this chapter. Along with basis theory, soybean quality was explained. Soybeans are desirable for their protein and more importantly their EAA content. Between the two, crude protein content is used by the market as an indicator of quality. Interestingly, the market does not have protein content requirements for transferring ownership of soybeans. According to Guinn (2002), soybean protein content levels can be tested by the Federal Grain Inspection Service (FGIS) upon request but are not required. Whether domestic or foreign, buyers that use soybeans as a source of feed for livestock want high levels of EAAs. If certain regions produce better quality soybeans in the form of EAAs then those regions will be targeted by buyers of soybeans in order to satisfy the needs of their feed rations. Though elevators have the option to test for protein, the majority of the testing is done by the buyer of soybeans at their processing facilities.

Quality is one of the four components of basis and plays an important role in commodity trading. Quality characteristics in soybeans vary spatially and may influence soybean demand. There is currently no value system in place to account for these varying levels of amino acid content in soybeans. Quality, in the form of EAA and crude protein, may have an implicit rather than explicit effect on origin basis values. This means that the market would still account for varying levels of quality even if it is not explicitly defined in the flat price.

CHAPTER 4. EMPIRICAL METHODS

4.1. Introduction

It is important for market participants to understand origin basis and its movements in order to minimize risk. As discussed in the literature forecasting or even interpreting origin basis movements can be complex due to the several explanatory factors that have been found to influence origin basis movements. Though research has been done to better forecast and explain basis movements there is still a large amount of uncertainty which opens the door for origin basis values to be studied further. Along with studying these explanatory variables there is a new frontier of research stemming from the importance of essential amino acids in soybeans being processed into soybean meal for feed.

Livestock, poultry, and other feed animals require dietary amino acids for good health, growth, and other survival needs. Nutritionists use soybean meal to satisfy a large amount of these amino acid dietary needs. This means there needs to be high amino acid content in the soybeans that are processed into soybean meal. Hertsgaard`s (2015) research shows Upper Midwest soybeans have superior levels of amino acids compared to soybeans grown elsewhere which make them more desirable in comparison to soybeans with lower amino acid content. Essential amino acids rather than crude protein being the desired quality variable in soybeans would be is positive news for soybean producers and other market participants involved with soybeans in North Dakota.

Though EAAs are the desired quality characteristic in soybeans market participants along the supply chain do not examine amino acid content in soybeans. Protein content can be examined upon request however, it is not a required standard practice. For this reason, it remains to be seen how amino acid content is factored into the market as a proxy of soybean value. The relationship between the explanatory and quality variables (in the form of amino acids and protein) on origin basis in North Dakota has not been determined.

There needs to be a better understanding of these market factors so that market participants involved with the soybean supply chain in North Dakota can interpret how significant certain explanatory variables are on origin basis values in the state. With North Dakota having been found to have low crude protein content and high essential amino acid content this research is of importance. A panel dataset comprised of time series and cross-sectional data is created that includes previously stated market factors.

4.2. Panel Regression Data

The data gathered to estimate the model is comprised of time series and cross-sectional data forming a panel data set. The length of the time series data is ten years and ranges from 2009 to 2018. The cross-sectional data is gathered across eight of the nine agricultural statistic districts defined by the USDA. The ten years of time-series data across the eight cross-sectional units form a data set of eighty observations. The model utilizes a wide variety of variables meaning that numerous sources are used to gather the data ranging from government entities to private servers.

North Dakota origin soybean basis, EAA, and protein values are gathered from eight of the nine North Dakota agriculture statistics districts. Analysis on the Southwest district is omitted from this study due to the lack of EAA data for that district. The districts, which are shown in Figure 4.1, were created by the United States Department of Agriculture and are used to report agricultural information among counties with similar agricultural characteristics. Daily basis values are gathered from one origin in each district using Bloomberg. Table 4.1 lists each origin in their corresponding district. The EAA and protein data for this study is gathered from the yearly North Dakota soybean quality reports published by the Northern Crops Institute which are funded by the North Dakota Soybean Council. Soybean samples are gathered from the Agricultural Statistics Districts in North Dakota and measure quality aspects such as protein, oil, and amino acid content on all the samples. Not all of the quality data measured in these reports are used in this thesis. As previously stated, this study examines the five most critical essential amino acids for livestock feed rations. The data provides one protein value and one value for each EAA per soybean crop year which in the United States runs from August 31st to September 1st.

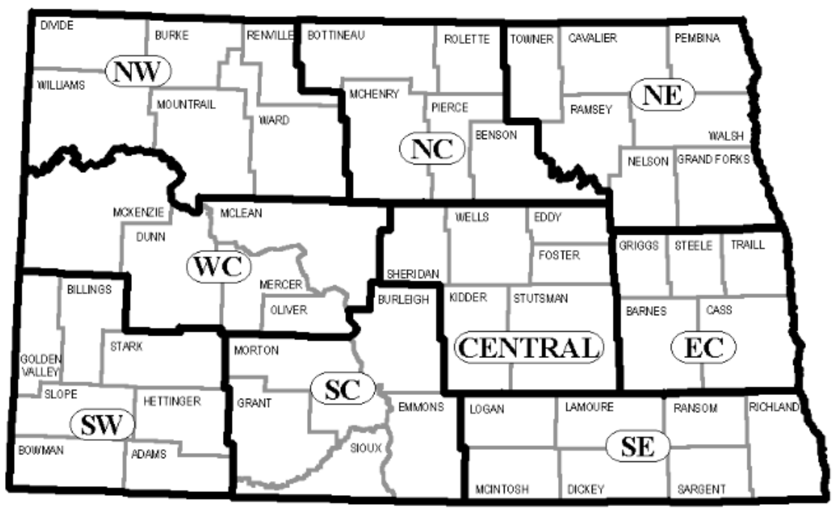

Figure 4.1. North Dakota Agricultural Statistics Districts

Table 4.1. Agricultural Statistic District Origins

District	Origin
Northeast	Cenex Harvest States Edmore, ND
North Central	BTR Farmers Leeds, ND
Northwest	Berthold Farmers Elevator Berthold, ND
East Central	Arthur Companies Page, ND
Central	Fessenden Cooperative Fessenden, ND
West Central	Wilton Farmers Union Wilton, ND
South-Central	South-Central Grain Hague, ND
Southeast	Farmers Elevator Hankinson, ND

Unlike the origin basis and quality values, the remaining data used in this study does not vary across the statistical districts. For the purposes of this study secondary car values represent shipping costs to origins in this study. Weekly Burlington Northern Santa Fe secondary car values are gathered using Thomson Reuters Eikon. Daily export basis values at the PNW are also gathered using Thomson Reuters. Daily soybean futures prices for the nearby and first deferred month are gathered using Thomson Reuters as well. These values are used to determine the futures spread between the nearby and deferred months over the complete time series.

4.3. Data Adjustments

All of the data in the panel dataset are arranged into quarterly and market year averages using Tableau. The data points in this study are averaged across market year and by quarter producing a total of five panel datasets. As discussed, the soybean crop year in the United States is from September 1st through August 31st. The quarter one (Q1) model averages values from September through November. This model will capture the effects of the variables on origin basis during soybean harvest in North Dakota. The quarter two (Q2) model averages values from December through February. This model represents the first post-harvest period. The quarter three (Q3) model averages values from March through May and is the second post-harvest period examined. The quarter four (Q4) model averages values from June through August and is the third post-harvest period. The market year (MY) model encompasses values from the entire soybean market year.

The origin basis time series history for the Central and South-Central districts had occasional missing values. Interpolation techniques are commonly needed to fill missing data points within a data set. The South Central and Central districts have missing values that are interpolated using k-nearest neighbors in X. According to Baretta and Santaniello (2016) k-nearest neighbors smoothing technique is a common variation of nearest neighbors smoothing that averages the k-closest values to the missing value.

4.4. Empirical Model

The models in this study are created to determine the relationship between quality characteristics along with explanatory variables on origin basis in North Dakota. Utilizing the panel dataset, the model developed in this study is written as:

$$B_{i,t} = \alpha_i + \beta_{1i} \bullet F1_{i,t} + \beta_{2i} \bullet F2_{i,t} + \beta_{3i} \bullet Pro_{i,t} + \beta_{4i} \bullet ExBasis_t \quad (4.1)$$

$$+ \beta_{5i} \bullet SCV_t + \beta_{6i} \bullet Sprd_t + \mu_{i,t}$$

Where “i” is the eight cross sections and “t” is the time series values in the form of market year or quarterly averages. Table 4.2 gives the variables used in the model as well as their description. The parenthesis represents the variable’s written form within the model.

Table 4.2. Model Variables

Variable	Description
Origin Basis (B)	The basis values from North Dakota origins.
Factor one (F1)	The first essential amino acid principal component.
Factor two (F2)	The second essential amino acid principal component.
Protein (Pro)	Soybean protein content.
Export Basis (ExBasis)	Soybean export basis at the Pacific North West export terminal.
Secondary Car Value (SCV)	Secondary car value variable charged to soybean rail shippers.
Futures Spread (Sprd)	Futures spread between nearby and deferred soybean contracts.

4.5. Principal Component Analysis

Principal components analysis (PCA) is a common method used to analyze multivariate data. PCA is applied in this study to sustain degrees of freedom as well as account for multicollinearity issues among variables. PCA involves reducing the number of highly correlated variables in a model while keeping the same amount of information in the form of principal components. According to Addinsoft (2020) PCA projects data from a larger dimensional space to a smaller dimension while retaining the maximum amount of information. The information they refer to is the variance of the data. Principal components are underlying factors that represent correlated variables. XLSTAT, a software by Addinsoft (2020), is utilized to run a PCA on the amino acid data for the five amino acid variables to identify principal components. The PCA produces five factor variables (principal components) that correspond with the five amino acid variables in the analysis. In PCA the first factor variable accounts for the largest amount of data variability as possible while the second factor variable accounts for as much variability left as possible and so on. If a large amount of the variability in the amino acid variables are explained in the first two factors, we are able to limit the representation of quality variables in the model into two quality variables.

The PCA gives a wide variety of information about the variables being analyzed. Squared cosign tables, factor correlation circles, and factor observations are a few of the results of a PCA that are used in this study. Squared cosines are what is used to interpret the amount of amino acid variability that is explained by the principal components. According to Addinsoft (2020), the squared cosines are used to determine the correct number of principal components needed to represent the data being tested.

4.6. Correlation Matrices

Correlation matrices are used to show correlation between variables by displaying correlation coefficients. A correlation matrix is produced using Pearson correlation in XLSTAT to identify correlation between the amino acids and protein in each district. Correlation coefficients define how much of a variable's variability is explained by another variable (Addinsoft 2020). According to Addinsoft (2020), Pearson correlation is the appropriate approach when working with continuous data and measures linear correlation between variables producing coefficients between negative one and one. P-values for the correlation matrix are also produced to show the level of significance of the correlations.

4.7. Model Specification

Model specification testing is necessary when constructing a regression model that uses a panel data set. Pooled OLS, fixed effects, and random effects models are three different models used for studying panel data. Pooled OLS requires panels that are pooled independently, because the data in this study contains eight cross-sectional panels observed from 2009-2018 pooled OLS will not be appropriate. This leaves a decision between using a fixed or random effects model.

The fixed effects model “fixes” or holds constant the model parameter’s effect on the dependent variable. Fixed effects models also are used if there is a possibility of omitted variable bias. According to Williams (2018) fixed effects models fix the variables to be time invariant so that the effects of the variables are the same across time. The random effects model, unlike the fixed effects model, retains model parameters as random variables. Random effects models do not control for omitted variable bias and assumes that these omitted variables are not correlated with the other explanatory variables.

Two tests are used to determine whether the panel data set requires a random or fixed effects model for estimation. The Breusch-Pagan and Hausman specification test are applied using the econometric software GNU Regression, Econometrics and Time-series Library (GRETL). The null hypothesis of the Breusch-Pagan Test is that the pooled OLS model is adequate. The alternative hypothesis is that the random effects model is appropriate. If the results of the Breusch-Pagan test produce small p-values less than 0.05 we reject the null hypothesis. The null hypothesis of the Hausman specification test is that the random effects model is preferred. The alternate hypothesis is that the fixed effects model should be used. If the results of the Hausman test produces a small p-value that is less than 0.05 you reject the null hypothesis and accept the alternative hypothesis that the fixed effects model is appropriate.

4.8. Summary

This chapter describes the methods from which the research is conducted. In this chapter a panel dataset is created which allows for the models to account for spatial variability among variables being tested throughout time. Adjustments to the data are made to produce a complete dataset. The model created in this chapter is regressed on the quarterly and market year data. PCA is applied to EAA values to identify principal components within the variables. The appropriate model specification tests were conducted to identify that using fixed effects is appropriate.

CHAPTER 5. RESULTS & ANALYSIS

5.1. Introduction

This section analyzes the results in the same chronological order as the testing is conducted. The first part of this section explains the results of the preliminary testing. The second section examines the empirical results of the model developed in this study and identifies variable significance in each model. The objectives of this study is to identify the influence quality and other explanatory variables have on origin basis in North Dakota. These variables are econometrically regressed using market year and quarterly averages of both the dependent and independent variables. The results of the model determine in which time of year amino acid content, protein content, and other explanatory variables influence origin basis the most.

As explained in the previous chapter, there are a number of preliminary tests that are needed to produce a well-defined central model that explains the relationship soybean quality and other explanatory variables have on origin soybean basis in North Dakota. The principal components analysis, variables characterization test, and model specification tests are conducted prior to forming the model developed in this study.

5.2. Overview of Results

Each model produces significant results that provide useful information to market participants in North Dakota. The quality variables are mostly significant in the second and third quarter models. This is because the quality characteristics are not known to the market until the post-harvest periods. Secondary car values which represent shipping costs are significant in every model which shows the importance of this variable on origin basis values. Secondary car values negatively influence origin basis values which aligns with the reviewed literature. Export basis values are also largely significant across every model besides quarter four. The futures spreads between the nearby and first deferred futures month did not have a significant influence on the dependent variable.

5.3. Data Adjustments

The PCA plots observations (data scores) in factors one and two. Figure 5.1 displays the amino acid observations from the PCA when ten years of data are analyzed from 2009-2018. Figure 5.2 shows the amino acid observations when 2009 is omitted from the data set. As shown, the amino acid variables for 2009 are outliers in the dataset compared to the rest of the observations the values are largely negatively loaded on factor one. When looking at the raw data, the amino acid values for 2009 are much lower than the values for the same amino acids in the rest of the data set. For these reasons, 2009 is omitted from this study so that there can be an accurate interpretation of the essential amino acid variables on our dependent variable. This leaves a time series of nine years ranging from 2010-2018.

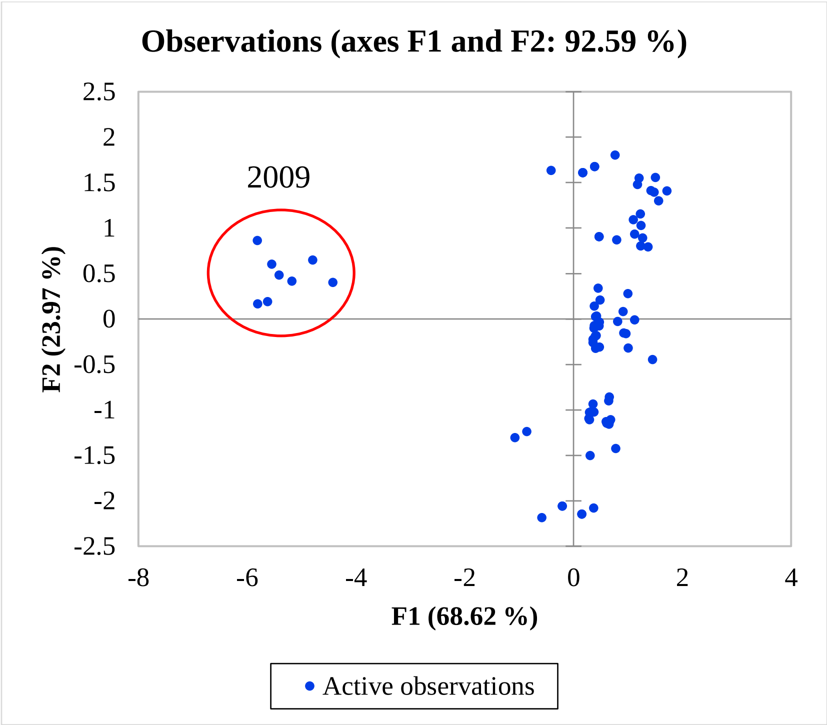

Figure 5.1. Ten Year Essential Amino Acid Factor Scores

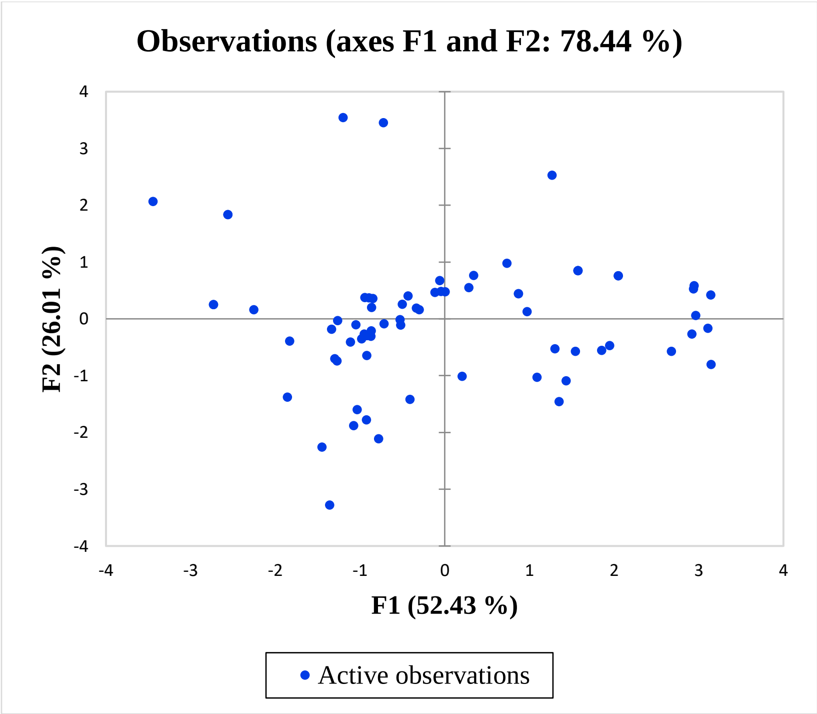

Figure 5.2. Nine Year Essential Amino Acid Factor Scores

5.4. PCA Results

The results of the PCA identify the squared cosigns of the factor loadings. According to Addinsoft (2020), the sum of squared cosines values across principal components is equal to one for each variable. The squared cosigns of the amino acids in each principal component are represented in table 5.4. As shown, five principal components are analyzed in the PCA and can be seen in the table. The bold values signify which factor describes the specific amino acids are most heavily loaded on.

Table 5.1. Essential Amino Acid PCA Squared Cosines

	F1	F2	F3	F4	F5
Cysteine	**0.780**	0.026	0.040	0.106	0.047
Lysine	**0.706**	0.013	0.161	0.119	0.000
Methionine	**0.405**	0.319	0.225	0.018	0.034
Threonine	0.020	**0.855**	0.078	0.015	0.032
Tryptophan	**0.710**	0.087	0.090	0.007	0.106

The principal components analysis of the EAA also produce a correlation circle. The correlation circle (or variables chart) shows the correlations between the components and the initial variables (Addinsoft 2020). The correlation circle shows the same results as the squared cosines in table 5.4. Figure (5.1) is a visual representation of the percentage each amino acid is on the factors with factor one along the x-axis and factor two along the y-axis. Factors further to the right are heavily loaded on factor one. Factors closer to the top are more heavily loaded on factor two. The correlation circle only represents the first two factors (principal components) because the large majority of EAA variability is explained by them.

As shown in the table all of the amino acids except for threonine are largely represented by factor one. Threonine is largely loaded on factor two with methionine being almost evenly split between the two factors. Factors one and two alone account for 78.44% of the variation among all amino acids in this study. While the remaining factors in the analysis are dropped, the first two factors are used in this study to represent the amino acids values. Factors one and two are able to accurately describe the amino acid values in the study by explaining over 75% of the variation. Having two quality factor variables rather than each individual amino acid in the model lessens the number of quality variables in the model while still allowing the amino acids to be accounted for.

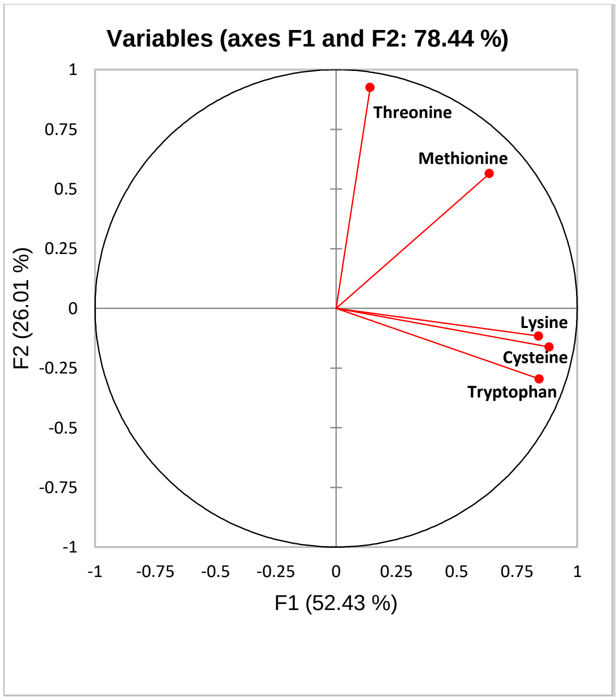

Figure 5.3. Correlation Circle

The results from the PCA produce factor loadings which represent the correlation between the EAA and the principle components. Table 5.2 shows the factor loadings for the five EAA. These factor loadings are used to derive the loading weighted coefficients and standard errors in the following chapter.

Table 5.2. PCA Factor Loadings

	F1	F2	F3	F4	F5
Cysteine	0.883	-0.162	-0.201	-0.326	-0.216
Lysine	0.840	-0.116	-0.401	0.345	0.009
Methionine	0.636	0.564	0.475	0.134	-0.183
Threonine	0.141	0.925	-0.280	-0.122	0.179
Tryptophan	0.843	-0.296	0.299	-0.083	0.325

5.5. Correlation Matrix Results

The Pearson correlation coefficients of the correlation matrices between protein and the amino acids by district are shown in table 5.3. As the table shows, the coefficients for the essential amino acids and protein content are low. Tryptophan and protein are negatively correlated in every district. Cysteine is positively correlated with protein in every district except the Northwest and West Central districts. Lysine, methionine, and threonine are positively correlated with protein in every district.

Table 5.3. Pearson Correlation Coefficients

	CENT	EC	NC	NW	NE	SC	SE	WC
					Protein			
Cysteine	0.120	0.045	0.042	-0.027	0.056	0.104	0.129	-0.153
Lysine	0.209	0.175	0.154	0.038	0.155	0.238	0.201	0.050
Methionine	0.229	0.194	0.206	0.055	0.222	0.216	0.183	0.089
Threonine	0.303	0.247	0.253	0.092	0.252	0.265	0.232	0.160
Tryptophan	-0.080	-0.159	-0.116	-0.146	-0.098	-0.041	-0.162	-0.150

P-values are also analyzed when correlation coefficients are identified. XLSTAT provides a p-value that corresponds with every correlation coefficient between the amino acids and protein. Table 5.4 shows the p-values associated with each correlation coefficient. The p-values in each district are well above the critical value of 0.05 indicating a failure to reject the null hypothesis that the true coefficient values in Table 5.3 are different from zero at the 95 percent confidence level.

Table 5.4. Pearson Correlation P-Values

	CENT	EC	NC	NW	NE	SC	SE	WC
					Protein			
Cysteine	0.758	0.909	0.914	0.946	0.886	0.791	0.742	0.694
Lysine	0.590	0.652	0.692	0.923	0.690	0.538	0.605	0.898
Methionine	0.553	0.616	0.594	0.888	0.565	0.576	0.638	0.820
Threonine	0.427	0.523	0.511	0.814	0.513	0.491	0.548	0.681
Tryptophan	0.838	0.683	0.767	0.707	0.801	0.916	0.678	0.701

5.6. Model Specification Results

The previous chapter explains the importance of model specification testing when determining whether a fixed or random effects model is appropriate for the panel data set. P-values are used to either reject the null hypothesis or fail to reject the null hypothesis of both tests. The results in the form of p-values for both the Hausman and Breusch-Pagan test on the market year and quarterly models are found in table 5.5.

The p-values from the Breusch-Pagan test are low (< 0.05) meaning we reject the null hypothesis that the pooled OLS model is appropriate and accept the alternative hypothesis of the random effects model. The p-values from the Hausman test are also low (< 0.05) meaning we reject the null hypothesis of a random effects model being the correct specification. By rejecting the null hypothesis of both the Breusch-Pagan and Hausman tests we are able to accept the alternate hypothesis of the Hausman tests which identifies the fixed effects model being appropriate. As shown in Table 5.5, quarter four is the only model in which the p-value is not less than 0.05. The results of the two tests consistently identify the fixed effects model to be appropriate.

Table 5.5. Model Specification P-Values

	Hausman	Breusch-Pagan
Market Year	0.0003	< 0.0001
Quarter 1	0.0111	< 0.0001
Quarter 2	< 0.0001	< 0.0001
Quarter 3	< 0.0001	< 0.0001
Quarter 4	0.0533	< 0.0001

Recall that fixed effects models are used when there is the possibility of omitted variable bias in a dataset. With the purpose of this study being to identify whether soybean quality and a handful of other explanatory variables influence origin soybean basis in North Dakota there are undoubtedly omitted variables. The results of the model specification testing align with this notion. Just like models need to be specified, data must also be analyzed to identify certain features.

5.7. Panel Regression Results

The model developed in this study is applied to market year and four quarterly panel data sets using GRETL. The panel regression results from the fixed effects models are shown in table 5.6. The coefficients, significance levels, standard errors, R-squared, and number of observations in each model are given. Each model has 58 degrees of freedom. The degrees of freedom are found by the formula "nt-n-k" where n is the number of cross-sections, t is the number of time periods, and k is the number of parameters in the model.

Table 5.6. Panel Regression Results

Independent Variable	Dependent Variable (Origin Basis)				
	MY	Q1	Q2	Q3	Q4
Factor 1	0.0084	0.0221	0.0345***	0.0249**	−0.0323
	(0.0113)	(0.0185)	(0.0083)	(0.0111)	(0.0267)
Factor 2	−0.0287	−0.0634**	−0.0394***	−0.0413**	0.0072
	(0.0188)	(0.0279)	(0.0113)	(0.0158)	(0.0500)
Protein	0.0150	0.0119	0.0202***	0.0396***	0.0116
	(0.0094)	(0.0100)	(0.0064)	(0.0085)	(0.0227)
PNW Export Basis	0.6838***	1.1877***	0.6671***	0.8227***	0.7890
	(0.2410)	(0.2550)	(0.2005)	(0.1612)	(0.6109)
Secondary Car Values	−0.0002***	−0.0003***	−0.0001***	−0.0003***	−0.0003*
	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0001)
Futures Spread	−0.0212	0.8785	−0.3467	−0.1203	0.1377
	(0.2965)	(0.6674)	(0.5483)	(0.1471)	(0.3498)
Constant	−2.1969***	−2.6081***	−2.2840***	−3.1585***	−2.2678*
	(0.5175)	(0.5132)	(0.4057)	(0.3878)	(1.2505)
Observations	72	72	72	72	72
R-Squared	57.10%	59.40%	83.30%	78.40%	13.90%

Standard errors in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

The Marketing Year (MY) model averages the variables across the entire market year from September 1st through August 31st. Factor one is insignificant and has a positive coefficient. Factor two is also insignificant and has a negative coefficient. Protein is not significant in the model and has a positive coefficient. This model had two highly significant independent variables. Export basis is also significant at the 1% level and has a positive coefficient. Secondary car values are highly significant at the 1% level and have a negative coefficient. Futures spreads are insignificant and have a negative coefficient.

The quarter one (Q1) model averages the variables between September 1st and November 30th. Recall that the quarter one model represents the harvest period. Factor one is insignificant in this model and has a positive coefficient. Factor two is significant at the 5% level and has a negative coefficient. Protein is insignificant and has a positive coefficient. Export basis is significant at the 1% level and has a positive coefficient. Secondary car values are significant at the 1% level and have a negative coefficient. Futures spreads are insignificant and have a positive influence on the dependent variable.

The quarter two (Q2) model averages variables between December 1st and February 28th. This is the first post-harvest period for soybeans in the state. Every variable besides futures spreads are significant in this model. Factor one is significant at the 1% level with a positive coefficient. Factor two is significant at the 1% level and unlike factor one it has a negative coefficient. Protein is significant at the 1% level and positively influences the dependent variable. The export basis variable has a strong, positive coefficient and is significant at the 1% level. Secondary car values are significant at the 1% level and has a negative coefficient. Futures spreads are insignificant with and negatively affect the dependent variable.

The quarter three (Q3) model averages the variables between March 1^{st} and May 31^{st}. This is the second post-harvest period. Factor one is significant at the 5% level in the quarter three model with a positive coefficient. The factor two variable is significant at the 5% level with a negative coefficient. Protein is significant at the 1% level with a positive coefficient. The export basis variable is significant at the 1% level with a positive coefficient. Secondary car values are significant at the 1% level in the quarter three model and have a negative coefficient. Futures spreads are insignificant in this model and have a negative coefficient

The quarter four (Q4) model averages the variables between June 1^{st} and August 31^{st}. This is the third post-harvest period. Factor one is insignificant in quarter four and has a positive coefficient. Factor two is insignificant and has a positive coefficient. Protein is also insignificant and positively influences origin basis. Export basis is not significant and has a positive coefficient. Secondary car values are the only significant variable in the quarter four model. Secondary car values have a negative coefficient and are significant at the 10% level.

Table 5.7 shows the loading weighted coefficients and standard errors of the EAA in the first two principal component variables. The coefficients are derived by multiplying the principal component coefficients (F1 and F2) with their corresponding EAA factor loadings and adding the values. The loading weighted standard errors are derived by taking the squared factor loading for the two principal components (F1 and F2) times the squared standard error for each principal component (F1 and F2) and adding those values followed by taking the square root of the summed value.

Table 5.7. Empirical Results

Amino Acid	MY	Q1	Q2	Q3	Q4
Cysteine	0.0121	0.0298*	0.0369***	0.0287***	-0.0297
	(0.0104)	(0.0170)	(0.0076)	(0.0101)	(0.0249)
Lysine	0.0104	0.0259	0.0336***	0.0257***	-0.0280
	(0.0097)	(0.0159)	(0.0071)	(0.0095)	(0.0232)
Methionine	-0.0109	-0.0217	-0.0003	-0.0075	-0.0165
	(0.0128)	(0.0197)	(0.0083)	(0.0114)	(0.0329)
Threonine	-0.0254	-0.0555**	-0.0316***	-0.0347**	0.0021
	(0.0175)	(0.0259)	(0.0105)	(0.0147)	(0.0464)
Tryptophan	0.0156	0.0374**	0.0407***	0.0332***	-0.0294
	(0.0110)	(0.0176)	(0.0078)	(0.0105)	(0.0269)

Cysteine, lysine, and tryptophan all have a positive effect on origin basis values in North Dakota in every model except quarter four. Cysteine is significant at the 10% level in the quarter one model with significance increasing to the 1% level in the quarter two and three models. Lysine is also significant at the 1% level in quarter two and quarter three models. Tryptophan is significant at the 5% level in quarter one. It is significant at the 1% level in the quarter two and quarter three models. Methionine has a negative effect on origin basis values in North Dakota and is not significant in any of the models. Threonine has a negative effect on the dependent variable in every model except quarter four. It is significant at the 5% level in the quarter one and quarter three models while being significant at the 1% level in quarter two.

5.8. Summary

The results of this study showed many variables to be significant on origin basis values in North Dakota. The essential amino acids represented in factor one are significant in the quarter two and three models. Factor one has a positive effect on origin basis values in every model except quarter four. Factor two is significant in the first three quarterly models and has a positive effect in every model except quarter four. Protein has a positive effect in every model. It was highly significant in the quarter two and three models. Export basis is highly significant in every model except quarter four. Export basis has a positive effect on origin basis across all models. Secondary car values are significant in every model having a negative effect on origin basis. Futures spreads are insignificant on origin basis values in North Dakota having a negative effect on origin basis in every model besides the quarter one and quarter four. The results provide market participants involved with soybeans in North Dakota more information about origin basis. Unwinding the EAA quality principal components allows for the influence of each individual EAA on the dependent variable to be shown giving market participants further knowledge about the relationship of soybean quality and origin basis values in North Dakota.

CHAPTER 6. CONCLUSION

6.1. Introduction

This thesis identifies soybean quality characteristics among other market factors to examine their influence on origin basis values in North Dakota. The results give market participants more information when it comes to mitigating basis risk in North Dakota. The following section reiterates the problems for which the research was conducted. This section also reviews the results of the study as well as the methods used to derive them. The contribution this thesis gives to literature surrounding origin soybean basis in North Dakota is discussed. Implications of the research as well as avenues for further research are also given.

6.2. Problem and Objectives

Origin basis is an essential market mechanism that serves multiple purposes for the commodity marketing industry. As described in chapter 2, one of the four main components of basis values is quality. In 2015, Hertsgaard found that EAA content rather than crude protein content is a more desirable quality characteristic. High EAA content of the five most critical essential amino acids in soybeans are a desired by nutritionists formulating livestock feed rations. The EAAs are crucial to the well-being and growth of feed animals because they are unable to synthesize these needed amino acids themselves.

Along with quality characteristics other market factors that have been found to influence origin basis values are analyzed in this study. Shipping costs in the form of secondary car values, export basis values, and futures spreads are studied to determine their influence on origin basis values in North Dakota. Basis forecasting models have been developed and continue to be studied as a form of basis risk management. These forecasting models typically analyze past basis values to forecast future ones and for this reason vary in accuracy due to changing market factors that are not examined in these models.

Understanding the market factors that have been found to influence basis values is crucial to being able to understand basis movements. Shipping costs are an example of a market factor that has been found to influence origin basis. Export basis values are another market factor that are significant in determining origin basis values. Origins analyze export basis values and the cost of shipping a commodity to the export terminal and adjust their basis accordingly to the movements of these market factors. Futures spreads are also considered to be an important market factor as they are an indicator of supply and demand for a commodity.

Identifying relationship North Dakota origin basis values have with soybean quality and other market factors are objectives of this thesis. The first objective of this thesis is to identify the influence soybean protein and essential amino acid content has on origin basis values in North Dakota. The second objective is to identify the influence certain market factors have on origin basis values in North Dakota. Understanding these relationships allows market participants to better understand and mitigate their basis risk.

6.3. Methodology and Model

A panel dataset is constructed to determine the relationship of soybean quality and other market factors on origin basis in North Dakota. Historical basis values from 2009-2018 are gathered from one origin in eight of the nine Agricultural Statistic Districts in North Dakota and averaged using Tableau. Five essential amino acids and protein content are the quality characteristics of soybeans for which data is gathered. Export basis, secondary car values, and futures spreads are also gathered. These values are averaged using Tableau and do not change across cross-sections. Four quarterly and one market year model are developed to identify the significance of the variables during this time period. A PCA ran on the amino acid variables to identify principal components. The results of the PCA found that over 75% of the variability in the five most critical essential amino acids is explained by the first two principal components which are retained as EAA factor variables. Model specification testing took place using the econometric software GRETL to distinguish between using a random or fixed effects model. The results of the Breusch-Pagan and Hausman Specification Test showed that the fixed effects model is appropriate for estimation.

6.4. Overview of Results

This thesis produced multiple significant results that give answers to the previously identified problems. The results of this study are particularly beneficial to market participants in North Dakota. The results show varying levels of significance across the models which gives information about the seasonality of basis and the factors that influence it. This section analyzes the results of the study by variable.

Literature has shown shipping costs to be significant in determining origin basis values. Hart and Olson (2017) found shipping costs to significantly influence origin basis values. The results of this study support those findings with secondary car values being highly significant across every model and having a negative effect on origin basis. Regardless of the time of year shipping costs in the form of secondary car values influence origin soybean basis in North Dakota. As discussed, when the cost of shipping a commodity changes, basis values are adjusted accordingly to cover these costs.

The relationship between major export terminal and origin basis values has been and continues to be studied. PNW basis values are analyzed in this thesis due to their importance to North Dakota origins as a destination for soybeans. Export basis values at the PNW export terminal are significant in every model besides quarter four (Q4). Export basis values at the PNW have a large positive effect on origin basis. The results show that North Dakota origin basis values are adjusted according to export basis movements.

Futures spreads are an indication for market participants to either store their commodity when spreads are wide or to sell when spreads are narrow. This inevitably adjusts demand for a commodity in the market. This theory cannot be proved in the results of this thesis. Futures spreads are not significant on the dependent variable in any model and have a negative effect on origin basis in North Dakota.

The results of the thesis found that EAA and protein are significant on origin basis values in North Dakota. The factor one, factor two, and protein variables were each significant in two or more of the quarterly models. The first principal component (F1) is significant on origin basis values in the second and third quarterly models. These models represent the first and second post-harvest period in North Dakota. In quarter one, soybean harvest is taking place and the quality characteristics are relatively unknown to the market which is why factor one is not yet significant to origin soybean basis. By quarter two and three the essential amino acid content is known and the quality starts to become significant on the origin values. Recall, cysteine, lysine, methionine, and tryptophan were heavily loaded on the first principal component (F1).

The second principal component (F2) has a significant negative effect on origin basis in the first, second, and third quarter models. Though significant in all three of these models, the significance increases in the post-harvest periods of the second and third quarter models. This is similar to the first principal component when the variable becomes more significant as the quality is known in the second quarter and beyond. Recall, threonine was heavily loaded on the second principal component (F2).

The previous chapter also showed the loading weighted coefficients of each EAA analyzed in this study which further showed the relationship between these soybean quality characteristics and origin basis values in North Dakota. Unwinding the principal components allows for the significance of each EAA to be determined. As shown in the previous chapter the majority of EAA become significant in the quarters one through three which reiterates the conclusion that these quality characteristics become significant in the post-harvest period.

Protein has a positive effect on origin basis in every model, however the variable does not become significant until quarters two and three. Like the other quality variables protein content of the crop is relatively unknown to market participants until post-harvest. Protein becomes significant on origin basis once the content is known to the market.

Basis values have been known to show seasonal patterns due to various market factors. The model results, when analyzed together, show seasonality effects of the variables on origin basis values. The quality characteristics usually do not become known to the market until the second quarter of the market year. This can be shown in table 5.6 as all of the quality variables become significant at the 1% level during this time.

6.5. Contributions to Literature

This thesis adds a new branch of literature to soybean basis values in North Dakota and potentially beyond. The relationship between the soybean quality characteristics and origin basis values had not been previously studied. Protein and EAA content of soybeans are significant on origin soybean basis values in North Dakota. These findings add to the literature surrounding the quality component of basis values that was discussed in the theoretical framework.

The results define the relationship between other variables in this study and origin basis values in North Dakota. Secondary car values, export basis, and futures spread values are analyzed due to their significance on origin basis in other studies. Each of these variables had unique results that add to the literature surrounding these variables and origin basis in North Dakota.

6.6. Implications

The research and supplemental results produced implications of this thesis. The first is that though EAA and protein are significant on origin basis values the market does not set requirements for these quality characteristics. Protein content can be tested by the FGIS upon request, but it is not required before the commodity can be transacted. The second implication is that though the EAA factor variables show that quality is significant on origin basis in North Dakota the coefficients can be difficult to interpret. The final implication is that there needs to be a greater understanding as to why individual EAAs have different influences on origin basis values.

6.7. Continued Research

Until now, the significance essential amino acid and protein content has on basis had not been determined in any capacity. Creating a new branch of research that has never been done before presents an opportunity for continued research. This study can be replicated using data and variables from another state. This would give origins and producers in another state more information on the variables that influence their origin basis. Another option would be to increase the study area to include multiple states in order to account for larger spatial differences among soybean quality and basis values.

Depending on the research interests, different quality aspects of soybeans could be analyzed for their influence on origin soybean basis values. For example, soybean oil content is an important quality characteristic for people associated with soybean crushing. Determining the relationship among soybean oil content and origin basis would be research that could branch from this thesis. Along with this option, the researcher could study the significance of oil content in on soybean basis in proximity to soybean crushing plants.

This study can be repeated using a different commodity. For example, Hard Red Spring Wheat (HRSW) is a widely raised throughout much of the upper Midwest. The independent variables in this study could be used to determine their influence on HRSW basis in North Dakota. If the study were to be replicated using a different commodity it would be important to identify and study the desirable quality characteristics of that commodity.

Knowing that individual amino acids influence origin basis values differently provides another avenue for continued research. Threonine and methionine negatively influence origin basis values while the rest of the EAA study positively influence basis values. Identifying what causes these amino acids to influence basis differently becomes important for market participants to understand. Trying to raise soybeans with amino acids that favorably impact basis values would become a goal of soybean producers across the state and beyond. This thesis introduces avenues for continued research around this topic.

REFERENCES

Addinsoft (2020). XLSTAT statistical and data analysis solution. New York, USA. https://www.xlstat.com.

Beretta, L., & Santaniello, A. (2016). Nearest Neighbor Imputation Algorithms: A Critical Evaluation. *BMC Medical Informatics and Decision Making*, *16*(3), 74. https://doi.org/10.1186/s12911-016-0318-z

Bullock, D. W., & Wilson, W. W. (Eds.). (2019). *Factors Influencing the Gulf and Pacific Northwest (PNW) Soybean Export Basis: An Exploratory Statistical Analysis*.

Department of Poultry Science. (n.d.). *Amino Acids*. Georgia State University. https://poultry.caes.uga.edu/extension/poultry-nutrition/soybeans/amino-acids.html

Gnu Regression, Econometrics and Time-series Library [Economic Software] (2020, March 5). Ramanathan, R., Cottrell, A., & Lucchetti, R. Retrieved from http://gretl.sourceforge.net/

Guinn, J. (2002). *Domestic Quality Standards and Trading Rules and Recommended Export Contract Specifications for U.S. Soybeans and Products* (p. 17). http://ussec.org/wp-content/uploads/2012/08/Guinn_Quality_Standards_Trading_Rules2002.pdf

Hart, C., & Olson, F. (2017). Analysis of Grain Basis Behavior During Transportation Disruptions and Development of Weekly Grain Basis Indicators for the USDA Grain Transportation Report. *CARD Staff Reports*. https://lib.dr.iastate.edu/card_staffreports/82

Hertsgaard, David Jorgen. (2015). *Costs and Risks of Testing and Blending for EAA in Soybeans* [Thesis, North Dakota State University]. https://library.ndsu.edu/ir/handle/10365/27885

Hertsgaard, D. J., Wilson, W. W., & Dahl, B. (2018). Costs and Risks of Testing and Blending for essential amino acids in soybeans. *Agribusiness*, *35*(2), 265–280. https://doi.org/10.1002/agr.21576

Lakkakula, P., & Wilson, W. (2019). *Origin and Destination Basis Inter-dependencies in Soybeans: A Panel Data Analysis*

Sanders, D. R., & Manfredo, M. R. (2006). Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Methods. *Journal of Agricultural and Applied Economics*, *38*(3), 513.

Skadberg, K., Wilson, W. W., Larsen, R., & Dahl, B. (2015). Spatial Competition, Arbitrage, and Risk in U.S. Soybeans. *Journal of Agricultural and Resource Economics*, *40*(3), 442–456. JSTOR.843.

Taylor, M. R., Dhuyvetter, K. C., & Kastens, T. L. (2006). Forecasting Crop Basis Using historical averages supplemented with current market information. *Journal of Agricultural and Resource Economics*, *31*(3), 549–567. JSTOR.

Thakur, M. & Hurburgh,, C. (2007, June). Quality of US Soybean Meal Compared to the Quality of Soybean Meal from Other Origins. Journal of the American Oil Chemists' Society. 84:835-

USDA/NASS 2018 State Agriculture Overview for North Dakota. (N.D.). Retrieved January 14, 2020, from https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=NORTH%20DAKOTA

USDA—National Agricultural Statistics Service—North Dakota District Map. (N.D.). Retrieved January 13, 2020, from https://www.nass.usda.gov/Statistics_by_State/North_Dakota/Publications/County_Estimates/nddistmp.php

Williams, R. (2018). *Panel data 4: Fixed Effects vs Random Effects Models* (p. 8). University of Notre Dame. https://www3.nd.edu/~rwilliam/stats3/panel04-fixedvsrandom.pdf

Wilson, W. W., & Dahl, B. (2011). Grain Pricing and Transportation: Dynamics and Changes in Markets. *Agribusiness*, *27*(4), 420–434. https://doi.org/10.1002/agr.20277

2018 North Dakota Soybean Quality Survey. (n.d.). Northern Crops Institute. Retrieved January 15, 2020, from https://www.northern-crops.com/whats-new/2019/1/31/2018-north-dakota-soybean-quality-survey