NCCC-134
APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Short-Term Dynamics and Structural Changes in the United States and Brazil Soybean Basis: Seasonality, Volatility, Structural Breaks and Information Flows

by

David W. Bullock, William W. Wilson,
and Prithviraj Lakkakula

Suggested citation format:

Bullock, D. W., W. W. Wilson, and P. Lakkakula. 2020. “Short-Term Dynamics and Structural Changes in the United States and Brazil Soybean Basis: Seasonality, Volatility, Structural Breaks and Information Flows.” Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. [http://www.farmdoc.illinois.edu/nccc134].

Short-Term Dynamics and Structural Changes in the United States and Brazil Soybean Basis: Seasonality, Volatility, Structural Breaks and Information Flows

David W. Bullock, William W. Wilson, and Prithviraj Lakkakula[1]

Paper presented at the NCCC-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management
St. Louis, Missouri, April 20-21, 2020

[1] Research Associate Professor, University Distinguished Professor and CHS Endowed Chair in Risk and Trading, and Research Assistant Professor respectively in the Department of Agribusiness and Applied Economics, North Dakota State University. Address all correspondence to David W. Bullock, 624 Barry Hall, 811 2nd Avenue North, Fargo, ND 58102 or david.w.bullock@ndsu.edu. This project was funded as a USDA/AMS Cooperative Agreement # 16-TMTSD-ND-0004.

Short-Term Dynamics and Structural Changes in the United States and Brazil Soybean Basis: Seasonality, Volatility, Structural Breaks and Information Flows

Recently, the United States - China trade dispute has emphasized the importance of Brazil as a major export competitor in the global soybean market. In this paper, we examine the time series characteristics of United States and Brazilian soybean basis markets for seasonality, changes in mean level, intermarket information flows, and other time series behavior. We specifically examined basis at 31 origins and 2 export locations in the United States and a primary export market in Brazil. The results strongly support the presence of analog seasonality indicating that seasonal patterns vary greatly from year-to-year at all locations. Time series intervention analysis indicates that the United States - China trade dispute had a significant lasting effect on the basis level in the Brazilian market but not in the United States. Granger causality analysis of information flows between the origin and export basis markets prior to and after the announcement of tariffs in the United States - China trade dispute shows a significant dampening effect upon the information flows between the markets following the announcement of tariffs. These results are useful in that they can provide guidance to market practitioners in modeling basis forecasts and also provide useful information regarding the impact of the recent United States – China dispute upon the behavior of these basis markets.

Key words: soybean basis, analog seasonality, Box-Jenkins time series, Granger causality

Introduction

Intense competition exists in the soybean markets, both domestically and internationally. Within the United States, there is rivalry between the two dominant ports, the United States Gulf (USG) and the Pacific Northwest (PNW), which are impacted by many random variables that change over time. A rivalry also exists between the United States and Brazil which has now emerged to be the dominant soybean exporting country. Brazilian soybean exports have been growing at a rapid rate, where logistics, quality and other factors are important, notwithstanding the recent interventions. Some of the important features of this competition include: spatially differentiated production; bi-modal (annual) shipping patterns from each country; multiple ports from each supplier; secondary markets for rail cars and barge rates that are volatile and seasonal; excessive, though improving, waiting time in Brazil causing accrual of demurrage costs, ocean shipping costs differentials; quality differentials, exchange rates, among others. The effects of these features are manifested in basis values at respective ports and origins. Ultimately, countries and traders compete based on the basis and, therefore, an understanding of the interrelationships and dynamics are critical to understanding these markets.

International competitiveness in soybean markets has evolved over the past couple decades. Traditionally, the United States was the dominant exporter and has a highly developed logistical system for grains and oilseeds. Over time, Brazil has expanded its production and exports. Brazil has had a lesser developed logistical system, mostly dominated by trucks shipping to the dominate ports at Santos and Paranagua. The effect of this was for a higher cost logistical system, and characterized by periodic and elongated wait times. Over time, this system has been improving with new ports, better interior shipping and the effect has been for reduced wait times.

The spatial competition is further exacerbated by the recent United States - China trade dispute and the resulting Phase 1 agreement. The result of the trade dispute was to: 1) impose a tariff on shipments from the United States, 2) increase exports from Brazil, and 3) reduce those from the United States. The resulting Phase I agreement states that China can expand its agricultural imports from the United States, but, these are subject to a number of conditions one of which is that purchases would have to be at "market competitive prices".

The law of one price is probably not adhered in these markets due in part to the evolving maturity of the marketing system, in addition to the seasonal characteristics of production and exports, as well as logistics and storage capacity. Instead, there has always been notable price differentials, frequently exceeding costs differences, and exacerbated by seasonality and lack of transparency. Over time these markets have become relatively more transparent.

The purpose of this study is to examine the time series characteristics of the United States and Brazilian soybean basis markets from a seasonality, mean heterogeneity, Box-Jenkins autoregressive-moving average, heteroscedasticity (ARCH), and Granger causality perspective. Of particular interest is the potential impact of the recent United States – China trade dispute upon these markets from a time series perspective. This paper makes several contributions to the literature on commodity market analysis. One is the introduction of statistical methods that have not been commonly used in agricultural economics research such as the use of the standard normal homogeneity test (SNHT) to test for shifts in the mean basis level over time, and the use of the SEATS procedure for identifying the appropriate Box-Jenkins time series model to use. These statistical methods provide agricultural economics and industry participants with additional tools that can be applied to enhance understanding of commodity market dynamics and have many potential applications elsewhere. Second, the paper provides a better understanding of spatial and dynamic interrelationships in international agricultural markets. Only a few previous studies have addressed any of these. The changes identified here are notable and provide motivation for additional studies. Finally, for analysts of international markets and competition, this study provides insights that are important for understanding competition.

Background and Previous Studies

Basis behavior and volatility at both origin and export markets has become important for commodity and risk analysts in making risk management decisions. This is particularly the case for soybeans although the concepts are similar for other international commodities such as corn and wheat.

Background

The United States grain marketing system adopted a number of notable changes over the past decades. Among them, one is the growth in soybean production in the United States with substantial growth to the northwest of traditional growing regions. Similarly, there has been like growth in Brazil as soybean production expanded north and westerly.

There have been several changes in the United States logistics system. One is the substantial growth in exports from the PNW ports, which in the case of soybeans, expanded from less than 5 million metric tons in 2000/01 to a peak of 20 million metric tons in 2016/17, and has since declined sharply. Other changes in logistics include adoption of forward shipping instruments (secondary rail markets), shuttle rail shipping, and quite massive investments in the country handling and rail infrastructure -- all of which lower marketing costs. Soybean logistics is also heavily dominated by the United States river system which provides low cost and competitive barge shipping. However, costs of these shipments have substantial volatility (intra-year and inter-year), and the river system needs upgrading. Finally, United States growers have added substantial on-farm storage in recent years.

Inter-port competition within the United States was analyzed by Skadberg et al. (2015). Export locations at the USG and PNW were included along with origin basis at a large number of interior origins. Shipping costs from each origin to destination included tariff rates, fuel service charges and secondary market values for rail shuttles. The model was specified as a spatial stochastic optimization model using copula distributions to determine the most likely spatial arbitrage opportunities. The study concluded that: 1) origins in the Upper Midwest had become highly dependent upon the PNW as a destination market, 2) arbitrage payoffs vary regionally, and 3) vertically integrated trading firms can capitalize on spatial-arbitrage payoffs.

The grain marketing system in Brazil is evolving and has a much lesser developed interior shipping system (e.g., rail and barges, in addition to roads and bridges) compared with the United States. The impact of this is relatively high interior shipping and handling costs in Brazil (USDA-AMS 2019). There is less extensive on-farm storage and, in addition, the port infrastructure in Brazil has not been well developed. The growth in soybean production has also resulted in longer-haul shipping. In response, there have been efforts to expand the roadways (e.g., BR163), rail lines and barge facilities to diversify the logistical network. Concurrently, the port infrastructure has also been expanding -- there currently are multiple ports either under development or slated for expansion. The observed changes in the logistical system have resulted in intense rivalry between the United States and Brazil, particularly in serving major importing countries such as China.

Another feature of United States – Brazil export competition is soybean quality. Generally, it has been observed that quality from Brazil exceeds that of the USG (Park and Hurburgh 2002; Thakur and Hurburgh 2007) and the quality at the USG exceeds that of the PNW (Hertsgaard, Wilson and Dahl 2019; Wilson, Dahl and Hertsgaard 2020). However, this is random or periodic, and as production in each country expands geographically, the spatial distribution of its quality is changing. Nevertheless, it is common for traders to impute a quality differential in bids for soybeans across origins which encourages varying forms of inter-port blending to meet buyer requirements.

Some of these changes occurred within our study period and are important to the results of this paper. Importantly, Brazil had been exporting for many years. However, beginning in the 2016/17 marketing year, Brazilian soybean production rose sharply from 85 to 120 million metric tons. Another change occurred during the 2013/14 marketing year when the United States had reduced rail car velocity, which increased United States export basis (Lakkakula and Wilson 2020). Brazilian vessel waiting times were reduced from 20 days to about 2-3 days and Brazilian interior shipping costs decreased by $1.00 - $1.50 per bushel during this time period.

Finally, traders in both markets are heavily dependent on the Chicago Mercantile Exchange (CME) soybean futures for hedging purchases and/or sales from each market. Though the CME plays a critical role in futures price discovery, the correlations to offshore prices are poor and have deteriorated. Partly due to this, it has given rise to proposals to develop an alternative futures specification to include in some way Brazil port delivery (Rennison and Meyer 2018; Almeida and Durisin 2019). In addition, there are efforts for more detailed 'market assessments' (such as from Platts and AgriCensus) of Brazilian basis values via the over-the-counter (OTC) markets.

Previous Studies on the Basis

Empirical analysis of the basis has a long history in agricultural commodity markets. Some of these analyze basis at delivery markets. Other studies have analyzed the basis at non-delivery markets (e.g., Taylor and Tomek 1984; Parcell 2000; Lara-Chavez and Alexander 2006). A number of studies analyzed international behavior of the basis. Zhang and Houston (2005) analyzed how soybean production in South America and futures volatility impacted the basis. They found that both of these variables had a negative impact on the basis. Tilley and Campbell (1988) analyzed the USG HRW basis. They found that the weekly basis was mostly explained by exports, free stocks, and the grain embargo.

Haigh and Bryant (2000) analyzed the effect of barge and ocean freight values in international grain markets using a Vector Error Correction GARCH-in-Mean model. Results indicated both barge and ocean price volatility influence grain prices. Barge price volatility's impact is greater on both grain prices and marketing margins compared with ocean price volatility. More recently, Lakkakula and Wilson (2020) analyzed the origin and export basis for soybeans in the United States, and factors impacting this interdependency. Results indicated that: 1) the origin and destination basis are determined simultaneously, 2) there are numerous logistical and export factors that affect both basis values, and 3) changes in shipping costs had a greater impact on the export basis when compared to the origin basis.

Bullock and Wilson (2020) examined the impact of fundamental factors upon the marketing-year average basis at the two major United States export markets (USG and PNW). They also introduced the concept of analog seasonality which was applied to the export basis for both markets. Results indicated that the primary factors impacting the average United States soybean export basis were the Brazilian basis level, competition from domestic origins, and the level of imports by China. The type of seasonal analog prevailing in the basis markets was primarily influenced by logistical conditions (i.e., railcar shortages, secondary car values, etc.) rather than market competition.

Time series models have been used in recent studies to examine properties of soybean basis at origin market locations (Dhuyvetter and Kastens 1998; Sanders and Manfredo 2006; Taylor, Dhuyvetter and Kastens 2006; Hatchett, Brorsen and Anderson 2010; Onel and Karali 2014; Lee and Brorsen 2017; Thompson et al. 2019). Additionally, other studies have combined time series with fundamental approaches to soybean basis forecasting (Jiang and Hayenga 1997; Parcell 2000; Zhang and Houston 2005).

There have been several studies that analyze commodity prices using time series techniques (Siami-Namini and Hudson 2017; Saghaian 2010). Here we summarize a few recent studies related to the applications of Granger causality as well as studies that are relevant to soybeans. Bradshaw and Orden (1990) used Granger causality tests to analyze whether the real trade weighted agricultural exchange rate granger causes real monthly prices and export sales of wheat, corn, and soybeans. The authors find the evidence of Granger causality from the exchange rate to export sales but the results are not definitive in the case of the direction of Granger causality to prices.

Using panel vector autoregression and Granger causality tests, Rezitis (2015) examined the relationship between crude oil prices, agricultural commodity prices, and fertilizer prices. Results of the study indicate that there is a bidirectional Granger causality between crude oil prices and international agricultural prices as well as between US exchange rates and international agricultural prices.

Lakkakula (2018) estimated causal relationships among five fertilizer prices using asymptotic Wald tests and bootstrap resampling techniques using Granger causality analysis. Among others, one important result indicated that the urea price Granger causes all other fertilizer prices, including muriate of potash, triple superphosphate, rock phosphate, and diammonium phosphate.

Data and Methodology

The data used in this study was comprised of weekly closing time series observations over the 1/4/2004 to 12/27/2019 time frame for the basis (cash minus futures) in two United States export markets (USG and PNW), one Brazilian export market (Santos), and 31 United States origin markets spread across 8 states (Ohio, Indiana, Illinois, Iowa, Minnesota, North Dakota, South Dakota, and Nebraska). Missing values in the dataset were filled using either linear interpolation or the nonlinear iterative partial least squares (NIPALS) procedure (Wold 1973) depending upon the nature of the missing data. For some of the analyses in this study, the weekly data were converted into monthly averages of the weekly closing values using *Tableau* (Tableau Software 2020). For deriving additive seasonal indices, the monthly basis data were converted into marketing year (September through August) by month format for the 2004/2005 through 2018/19 marketing years.

Table 1 (Appendix A) lists the 31 United States origin basis markets and the electronic sources for the cash price data which were obtained from either *ProphetX* (Data Transmission Network 2020) or *Eikon* (Refinitiv 2020). These values were converted into nearby basis values by subtracting the nearby soybean futures price obtained from *Eikon* (code SC1). Basis data for the

United States export markets (USG and PNW) was from TradeWest Brokerage and are based upon CIF values. For Santos, the export prices were from *Eikon* (code: SB-FOBEXS-C1) and are FOB values per metric ton. These were converted from metric tons to bushels assuming a standard soybean test weight of 60 pounds per bushel.

Figure 1 (Appendix B) shows the 31 interior locations geographically along with the long-term average basis values using a color range. The average basis value declines when moving from East to West across the map. The Illinois, Indiana, and Ohio locations are mostly at a premium to the futures and represent mostly barge loading facilities on the Ohio, Illinois, and Mississippi Rivers. The average basis also declines along the Mississippi River when moving from South to North.

Figures 2 and 3 in Appendix B show monthly average soybean basis values for the USG and Alden, Iowa respectively along with the 12-month moving averages. Note that the moving averages from both graphs show the same basic long-term pattern which also holds across all of the United States locations in this study. A general weakening trend is observed through early 2008. This is followed by a strengthening trend through the beginning of the 2009/10 marketing year (Sept 2009). This is followed by a period of stability through the beginning of the 2012/13 marketing year when there was a sharp increase in the basis through the beginning of the 2013/14 marketing year. A brief period of stability is followed by a general decline in basis values beginning in mid-2015 through the end of 2019.

Figure 4 shows the Santos basis over the same time period with the 12-month moving average. The moving average pattern differs substantially from the observed United States pattern. Also, the Santos basis exhibits a much higher level of volatility when compared to the USG and PNW basis.

Basis Seasonal Analogs

For examining seasonality, the analog methodology presented in Bullock and Wilson (2020) was used on the monthly average time series. First, each of the data series were converted into a marketing year by month. Next, the additive seasonal indices were calculated by taking each monthly basis value and subtracting the corresponding marketing year average. To derive the seasonal analogs, the marketing years were clustered using the agglomerative hierarchal clustering (AHC) algorithm (Ward 1963) based upon dissimilarity by Euclidian distance with Ward's agglomeration and minimum-entropy (Shannon 1948) to determine the cutoff point on the dendrogram. From this procedure, a set of unique seasonal analogs was derived for each location represented by clusters of marketing years. These clusters are represented by an integer label for each marketing year by location with the initial marketing year (2004/05) always starting as '1' with the next differing marketing year beginning analog '2' and so forth.

The seasonal analog clusters were then visually examined to determine any commonalities between locations. Additionally, the correlation of seasonal analogs across locations was roughly derived by applying an extra layer of clustering to the analog labels arranged in a marketing year by location matrix. This clustering was done using AHC with similarity based upon a co-occurrence matrix using unweighted pair-group averaging for the agglomeration method with minimum entropy used to determine the dendrogram cutoff. An identical procedure was applied to the matrix for grouping similar marketing years across locations. All of the seasonal clustering in this study was done using the *XLStat* (Addinsoft 2020) computer software.

Homogeneity Tests of Mean Basis Levels

To determine if recent events, such as the emergence of Brazil as an export competitor and/or the recent trade dispute with China, have led to shifts in the average basis level for each location, a time series homogeneity test proposed by (Alexandersson 1986) called a standard normal homogeneity test (SNHT) was applied to each data series. The SNHT test was applied to the standardized normal values [$z_t = (x_t - \mu)/\sigma$] of the time series data. The null hypothesis (H0) of the test is that all z_t ($t = 1,\ldots,T$) follow a standard normal [N(0,1)] distribution. The alternate hypothesis (Ha) is that there exists a breakpoint v on the $1,\ldots,T$ interval where the z_t follow an $N(\mu_1, 1)$ distribution for $t = 1,\ldots,v$ and follow a $N(\mu_2, 1)$ distribution for $t = v+1,\ldots,T$ where $\mu_1 \neq \mu_2$. The test statistic (T_0) for the SNHT test is defined by:

(1) $$T_0 = \max_{1 \le v \le T}[v \cdot \bar{z}_1^2 + (T - v)\bar{z}_2^2],$$

where

$$\bar{z}_1 = \frac{1}{v}\sum_{t=1}^{v} z_t, \text{ and}$$

$$\bar{z}_2 = \frac{1}{T-v}\sum_{t=v+1}^{T} z_t .$$

The statistical distribution and p-value for the T_0 statistic was derived using Monte Carlo methods. In this study, the *XLStat* (Addinsoft 2020) statistical software was used to conduct the SNHT test on the monthly average time series for each location. To limit the testing to the previously mentioned recent events, the test was applied only to monthly data starting with the beginning of the 2008/09 marketing year (September 2008) through the end of the dataset (December 2019).

Box-Jenkins Time Series and Intervention Modeling

To determine the individual time series characteristics of each data series, a Box-Jenkins autoregressive integrated moving-average model (Box and Jenkins 1970) with a seasonal component (SARIMA) was fit to the monthly observations for each location. Normally, the identification of a Box-Jenkins time series model is an iterative process that follows the steps of (1) model identification, (2) model estimation, and (3) diagnostic testing. The model identification phase typically involves visual examination of the autocorrelation (acf) and partial autocorrelation (pacf) functions of the time series along with stationarity statistical tests. A candidate model is identified and estimated. Then the residuals of the estimated model are examined to determine whether they statistical fit a pattern of white noise (i.e., no significant spikes in the acf and pacf based upon statistics such as the Ljung-Box).

There have been some attempts to automate the Box-Jenkins model identification process. Most of these have been based upon estimating a standard set of models and choosing the best model based upon a diagnostic statistic or information criterion. An example of this would be the *Bestfit* time series model estimation procedure that is part of the *@Risk* Monte Carlo simulation software package (Palisade Software 2018). However, the seasonal capabilities of Bestfit are limited to pre-application of either seasonal differencing or additive seasonal indexing of the data which preclude the identification and estimation of true Box-Jenkins SARIMA models.

An automated procedure proposed by Gómez and Maravall (2001) called *Signal Extraction in ARIMA Time Series* (SEATS) was used to identify both the regular (ARIMA) and seasonal (SARIMA) Box-Jenkins time series components for each location's basis time series (monthly average). The SEATS procedure first determines the optimal level of regular and seasonal differencing through the estimation of unit roots using an iterative process of estimating SARIMA (1,1) x (1,1) on the differenced series. The optimal level of differencing is determined when the estimated model coefficients no longer produce unit roots. After the level of differencing is identified, the SEATS procedure identifies the regular and seasonal ARMA orders of the differenced series using a penalty function approach based upon the Bayesian Information Criterion (BIC) and first proposed by Hannan and Rissanen (1982).

The SEATS procedure that is part of the *X-13ARIMA-SEATS* (Monsell et al. 2013) seasonal adjustment software was used (with logarithmic transform and outlier detection set to 'off') to identify the level of differencing and order of the SARIMA model for each series. Then, the identified models were estimated using the 'arima' procedure in the *GNU Regression, Econometric and Time-series Library* (GRETL) econometrics software program (Cottrell and Lucchetti 2019). To test for time-varying volatility, an ARCH LM test (12 lags) was applied to the residuals from each fitted time series model. In addition, two simple fixed-effect intervention models were estimated for each time series by adding a dummy variable to the estimated model and testing for statistical significance of the dummy variable coefficient. The dummy variable (D1) represents the time period corresponding to and following the announcement of United States tariffs (March 18, 2018) on the import of Chinese goods and is equal to '1' between March 2018 and December 2019 and '0' otherwise.

Granger Causality and Information Flows

Granger Causality (Granger 1969) was used to determine information flows among multiple basis markets. The tests were run on the weekly time series. In order to simplify the analyses, the origin market locations were aggregated into two separate regions: (1) the Eastern Corn Belt comprising locations in Illinois, Indiana, and Ohio; and (2) the Western Corn Belt comprising Iowa, Minnesota, Nebraska, South Dakota, and North Dakota. The regional basis values were calculated as the simple arithmetic average of all locations within the region.

To examine the impact of the recent China trade dispute upon information flows, the weekly time series was divided into two periods: (1) the weekly data prior to the announcement of tariffs on 3/18/2018 which ran from 1/2/2004 through 3/16/2018, and (2) the data following the announcement which ran from 3/23/2018 through 12/27/2019. The Granger Causality tests were run on each dataset separately and the causal flows compared between the two periods. It is important that although the tariffs were announced on 3/18/2018, they did not go into effect until the following July.

To maintain comparable models between the two time periods, the same lag length was maintained between the two time periods and determined based upon application of the Hannan-Quinn Information Criterion (HQIC; Hannan and Quinn 1979) to each time period and retaining the longer of the two identified lag lengths. The HQIC was used because it provides a good compromise between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) which can often give conflicting results. For the pre-announcement time period, the HQIC identified an optimal lag length of three weeks while the post-announcement period had an optimal lag length of one week. Therefore, a lag length of 3 weeks was used in applying Granger Causality tests for both time periods. The lag lengths were determined using the 'var' command with the '---lagselect' option in GRETL with a max lag length of 24 weeks for the pre-announcement period and 6 weeks for the post-announcement period.

The presence of unit roots was determined using KPSS (Kwiatkowski et al. 1992) test to each individual time series and time period. The results of the tests indicated that all of the series had one unit root with the exception of the USG series in the post-announcement period which had no unit root. Due to the presence of unit roots, a test (TY) proposed by Toda and Yamamoto (1995) was used to test for Granger Causality between basis pairs.

The TY test basically involves estimating a vector autoregression model between each pair (x, y) of series in levels with lag length equal to $l + d$ with l equal to the optimal identified lag length and d equal to the maximum number of unit roots between the two series. To determine if series y "Granger causes" series x, a Wald chi-square test (degrees of freedom equal to l) is applied to the null hypothesis that only the first l lags of the series y coefficients in the series x equation are equal to zero. If the null hypothesis is rejected, then Granger causality is established from series y to series x. The same test is applied to the first l lags of series x coefficients in the series y equation to establish whether x causes y. Causality with feedback is established between x and y if each series causes the other. The causality results are further verified by application of Johansen cointegration tests (Johansen 1988; Johansen 1991) to determine if a cointegrating vector exists between the two series.

In this study, the TY test was applied to each pair of series (x, y) by estimating the VAR regression equations manually using the 'ols' command in GRETL with the lag length set to 4 weeks (i.e., 3-week optimal lag length plus 1 week to account for unit root). Causality was then tested by applying the GRETL 'omit' command on the first 3 lagged coefficients of the causing variable with the '—chi-square' flag activated. For application of the pairwise Johansen tests, the GRETL 'coint2' command was used with lag length set to 4 weeks.

Results

The estimation and statistical test results are presented in tabular (Appendix A) and graphical (Appendix B) format where appropriate. Each result is followed by a discussion of the key implications.

Seasonal Analog Analysis

A summary of the seasonal analog groupings (based upon application of AHC clustering using Euclidian distance dissimilarity) and the count of analogs by location is contained in Table 2. The table is formatted in a location (row) by marketing year (column) format. The numbers in the table indicate the seasonal analog identifier number. The numbers begin with the first marketing year as '1' and then subsequently assign a different integer in ascending order as a marketing year begins a new analog cluster. The analogs that comprised a single marketing year are highlighted in bold italics. These single year analogs are typically referred to as *outliers* and are generally characterized by a greater level of volatility when compared to the multiple year analogs.

For a majority of locations, the 2012/13 and 2013/14 marketing years were characterized by these single-year analogs. Bullock and Wilson (2020) indicated that these two marketing years were highlighted by logistical issues. This was primarily due to the confluence of increased demand for rail transportation from other sectors (coal, oil, intermodal, sand, and gravel), record 2013/14 corn production, sizable United States soybean and wheat crops, record Canadian wheat production along with severe winter weather which compounded existing rail capacity constraints (USDA-OCE and USDA-AMS 2015). Also, it was notable that the marketing years overlapping with the China trade dispute (2017/18 and 2018/19) were grouped into regular seasonal analog patterns for all locations with the exception of Santos which had a single-year analog pattern in 2018/19.

Figure 5 shows a histogram of the number of locations by number of analogs with a statistical summary of the number of analogs by location. Over a third (13 out of 34 or 38.2%) of the locations had just the minimum of 3 analogs. An additional quarter (9 out of 34 or 26.5%) of the locations had 4 analogs while the remaining 12 locations (35.3%) had 5 or greater analogs with just one location each in the top two categories (7 and 8 analogs). The mean number of analogs per location was just over 4 (4.24) while the Pearson skewness coefficient indicates the distribution is skewed to the upward side. These results indicate that there were many analogues and they varied across origins, and through time.

A plot of the three seasonal analogs for the PNW export market is illustrated in Figure 6. Analog 1 exhibits the prevailing seasonal pattern (11 out of 15 years) which is mostly a sideways pattern with slightly above average basis values during the 1st half of marketing year followed by below average basis values in the 2nd half of the marketing year. Analog 2 (3 out of 15 years) shows a general strengthening pattern that occurs in years with tightening carryout-to-use ratios (2007/08, 2011/12, and 2013/14) brought on by substantial increases in Chinese soybean imports. Analog 3 is a single-year (2014/15) outlier reflective of transportation disruptions.

Seasonal analogs for Bayard, Iowa are illustrated in Figure 7. Analog 1 (8 out of 15 years) is the normal sideways pattern in the basis. Analog 2 (5 out of 15 years) is a flat basis pattern with a strengthening basis during the final 4 months and is reflective of generally strong export demand. Analog 3 is a single-year (2012/13) outlier reflective of the severe drought in the summer of 2012. Analog 4 is a single-year (2013/14) outlier reflective of transportation disruptions.

Figure 8 shows a plot of the seasonal analogs for Santos, Brazil. Analog 1 (11 out of 15 years) is the normal sideways pattern with slightly stronger basis in the 1st half of the United States marketing year and a slightly lower basis in the 2nd half of the marketing year. Analog 2 is a single-year (2011/12) analog showing a volatile pattern with a complete collapse of the basis in the 2nd half of the marketing year. This collapse was a sharp decrease in exports after the peak-month of May of that marketing year. Analog 3 has two consecutive years (2012/13 and 2013/14) with a pattern reflective of the normal seasonal pattern (Analog 1) but with a much higher level of volatility. These years correspond to the emergence of Brazil as a major competitor on the international market. Analog 4 is a single-year (2018/19) outlier with a volatile pattern and is likely reflective of the extreme volatility brought on by the United States - China trade dispute.

Table 3 shows the result of applying AHC clustering (based upon co-occurrence similarity) to the locations based upon analog identifiers by marketing year. The clustering procedure placed the locations into 4 distinct groupings. The groupings are generally along the lines of the number of analogs with the first group containing the most locations (21 or 61.8%) with the fewest number of analogs (3, 4, or 5). The last two groupings contain only two locations each with the largest number of analogs (6, 7, or 8). Table 4 shows the co-occurrence clustering of the marketing years based upon analog identifiers by location. Almost all (13 out of 15) of the marketing years fall into the first grouping. The next two groupings contain single-year outliers for the 2012/13 and 2013/14 marketing years which reflect the years with high transportation costs and logistical difficulties.

Overall, two observations can be made regarding the seasonal analog analysis. First, the results indicate that seasonality is not homogenous across locations and time. The results clearly show that analog seasonality exists not only for the export markets (as shown in Bullock and Wilson 2020) but also can be extended to origin and international export (Santos, Brazil) markets. The total number of analogs varies widely across locations; however, a majority of the locations show volatile outlier analogs in marketing years (2012/13, 2013/14, and 2014/15) where there were logistical difficulties. Second, it appears that the recent United States - China trade dispute had little, if any, impact upon basis seasonality with 2018/19 grouped into the normal seasonal analog for all of the locations with the noted exception of Santos, Brazil which has it as an outlier analog.

Homogeneity Tests of Mean Basis Levels

A summary of the Standard Normal Homogeneity Test (SNHT) across all locations covering the monthly time period from the beginning of the 2008/09 marketing year (September 2008) through the end of the dataset (December 2019) is illustrated in Table 5. The results show breakpoints in the mean basis level ranging from October 2010 to September 2016. The T_0 statistics indicate statistically significant differences in the mean values at the 95% confidence level with the exception of five locations (all located in Illinois, Indiana, and Ohio). The origin locations with the early breakpoints (November 2011 or earlier) had an increase in the mean basis from the before break to after time windows. These locations are also exclusively located in the states east of the Mississippi River (Illinois, Indiana, and Ohio). Only two locations east of the Mississippi River (Mount Vernon, IN; and Nauvoo, IL) have late (Sept 2016) breakpoints with a negative change in the basis.

For origin locations west of the Mississippi River (Iowa, Minnesota, Nebraska, North Dakota, and South Dakota), the breakpoints occurred in November 2013 or later with statistically significant (95% or better) negative shifts in the mean basis level. The two United States export locations (USG and PNW) both had later breakpoints and statistically significant negative shifts in their mean basis levels. Santos, Brazil, on the other hand, had a later (August 2014) breakpoint with a statistically significant upward shift in the mean basis value.

Overall, two observations can be made from these results. First, the increase in the mean Santos, Brazil basis in August 2014 lead a subsequent decline in the PNW (December 2014) and USG (September 2016) mean basis values. Figures 9 and 10 show the SNHT test results graphically for the PNW and Santos respectively. Second, these results indicate that the emergence of Brazil as a major export competitor had a larger and more significant negative impact on United States average basis levels when compared to the more recent United States – China trade dispute.

Box-Jenkins Time Series and Intervention Modeling

The data was analyzed to determine the form of the time series characterization. The results of the Box-Jenkins time series identification using the SEATS procedure along with the results of the ARCH and China trade dispute intervention analysis are illustrated in Table 6. The most commonly identified Box-Jenkins model (monthly data) was the SARIMA (1, 0, 0) x (1, 0, 0)[2] which was identified in 13 (38.2%) of the locations. This was followed by the SARIMA (2, 0, 2) x (1, 0, 0) which was identified in 5 locations (14.7%) and the SARIMA (1, 0, 1) x (1, 0, 0) which was identified in 4 locations (11.8%). Only two locations (Finley, ND and PNW) had identified models without a seasonal component.

The Lagrange Multiplier (LM) test statistics supported ARCH behavior at the 95% confidence level in the time series residuals for the United States origin locations. The PNW had support (p-value of 0.0787) for ARCH behavior at the 90% confidence level. The intervention dummy variable for the China trade dispute announcement (D1) was mostly negative but statistically insignificant (at 95% level) for the United States locations. However, D1 was both strongly positive and significant for Santos, Brazil.

From these results, two major observations can be made. First, ARCH time series behavior is strongly present in the origin basis series but not for the export markets. Second, the China trade dispute had a statistically significant and strongly positive impact upon the mean basis level for Santos, Brazil while having no significant impact upon United States export and origin basis values.

[2] The Box-Jenkins SARIMA notation has the regular time series part in he first set of braces followed by the seasonal part in the second set. The terms in each part are listed as number of autoregressive lags, degree of differencing, and number of moving average lags. So, for example, a SARIMA (1,0,0) x (1,0,0) is a Box-Jenkins model with 1 regular and 1 seasonal autoregressive lag with no differencing or moving average terms.

Granger Causality and Information Flows

Finally, Granger causality tests were used to characterize the information flows among these markets. The Toda-Yamamoto Wald chi-square statistics for the pre-China trade dispute period (prior to March 18, 2018) are illustrated in Table 7. The causality relationships implied by these statistics are summarized geographically in Figure 11. Application of the Johansen cointegration tests supported the chi-squared test results.

These results indicate that these markets were highly interdependent with multiple feedback loops between the three export markets (PNW, USG, and Santos) and the two interior regions (Western Corn Belt, Eastern Corn Belt) prior to the tariff announcement. The primary integrator for all of the basis markets was the USG, which was strongly caused, from a Granger perspective, by all of the other markets. The information from the USG was strongly reflected in both the PNW and Santos markets.

For the post-China trade dispute period (weeks following March 18, 2018 through end of 2019), the Toda-Yamamoto Wald chi-square statistics are illustrated in Table 8 and summarized geographically in Figure 12. The Johansen cointegration tests supported the chi-squared results. The results indicated a substantial decline in the interdependency among the basis markets when compared to the pre-announcement period. The PNW and USG export markets became completely exogenous to the system while Santos was 'Granger caused' by the United States interior markets with no feedback loop present.

These results indicate that while there was a large degree of interdependency in the United States and Brazilian basis market, the onset of the China trade dispute with the United States had an important dampening effect on the information flows between these markets. Prior to the dispute, basis values between the various markets were likely influenced by the standard basis fundamentals such as transportation cost differentials, market competition, and seasonal influences. Subsequent to the announcement of tariffs, these interrelationships likely broke down as Brazil became the primary export market with the PNW and USG becoming secondary in terms of competition with little export volume flowing through either port. With the disfunction of the United States export markets, Santos (Brazil) mainly looked to the United States interior markets (which were still functioning) for information on these basis fundamentals.

Summary and Conclusions

The purpose of this study was to analyze the short-term dynamics of the United States origin and export, and Brazilian export basis values using time series methods. Specifically, we analyze seasonality, mean level homogeneity, Box-Jenkins time series identification and estimation, analysis of time varying volatility, and intervention analysis of the effects of the recent United States – China trade dispute, in addition to Granger causality among spatially separated markets for the time period preceding and following the onset of announced tariffs in the United States – China trade dispute.

Application of seasonal analog analysis using agglomerative hierarchal clustering (AHC) resulted in the identification of between 3 and 8 unique seasonal analogs for each of the 34 locations examined. A majority of the locations had at least one single-year (outlier) analog which mostly reflected marketing years characterized by logistical issues. Many locations also had at least one dominant analog that comprised a majority of the marketing years and was reflective of a relatively flat basis pattern across the marketing year. The results indicate that seasonality is widely variable from year-to-year and across locations. This has implications regarding the hazard of using annual moving average methods to forecast seasonal basis.

Application of the SNHT mean homogeneity test to the individual series resulted in identification of variable breakpoints in the United States locations that all occurred well before the beginning of the United States – China trade dispute in March 2018. Many of the identified breakpoint months for the United States locations corresponded with the emergence of Brazil as a major export competitor in the 2013 to 2016 time window. In almost all of the United States locations (particularly west of the Mississippi River and at the export locations), the mean basis fell significantly (35 to 60 cents per bushel) while the Brazilian (Santos) mean basis increased by 70 cents per bushel over the same time period.

Identification of seasonal Box-Jenkins time series (SARIMA) models was done for each series using the SEATS automated procedure. The results identified 11 different time series model specifications across the 34 locations with the SARIMA (1, 0, 0) x (1, 0, 0) as the most common (38% of locations). Lagrange Multiplier (LM) test results indicated the strong presence of ARCH behavior in the United States origin locations but failed to support ARCH behavior in the United States and Brazil export locations. To examine the potential impact of the United States – China trade dispute, a dummy variable representing permanent effects commencing with the March 18, 2018 announcement of tariffs was added to the identified SARIMA models to estimate time series intervention models. The dummy variable coefficient was only statistically significant at the Santos, Brazil location and was strongly positive indicating that the United States – China trade dispute had a positive and lasting impact upon only the Brazilian export basis.

Application of Granger causality tests for the time period preceding the announcement of tariffs (March 17, 2018) in the recent China – United States trade dispute strongly supported extensive bidirectional causality across all of the markets. The dominant information aggregator during this period appears to be the USG export market. In the period following the tariff announcement, the results failed to support Granger causality with the exception of unidirectional causal flows from the United States origin markets to Santos, Brazil. Both of the United States export locations became completely exogenous to the system in this subsequent time period.

Taken together these results have several implications. One is the significant structural shift resulting in increases in the Brazil basis, and decreases of those in the United States (over 100c/b) is very notable. This differential is ultimately transmitted to growers as changes in relative prices. That these relationships are changing creates challenges for growers and for commodity marketing firms in making trading and risk management decisions, as well as decisions about investment in grain marketing infrastructure. Further, these changing relationships have a greater incidence at locations in the Western Corn Belt.

Second, these results show that the effect of the Chinese import tariffs are clear, resulting in a significant impact on the Santos basis. This impact has an important effect on traders and growers in each country. The extent that these changes are transitory, or if the time series of the bases revert to the prior regime remains to be seen. Of particular importance is that there is a clause in the United States-China Phase 1 agreement which states that imports would be made at "market competitive prices." Of course, this term is not defined. As these results indicate, the dynamic interdependence among these prices suggests that assessing whether prices are at 'market competitive' levels, will be difficult and subject to seasonal, and prospective structural shifts, notwithstanding their non-stationarity.

Finally, the results have implications for market analysts, as well as firms and organizations involved in price discovery and assessment. Prior to the Chinese imposition of tariffs, these basis markets were highly interdependent with numerous unidirectional and bi-directional feedback loops resulting in observed values, and the USG was the dominant aggregator of information. This is to be expected. However, there have been significant structural changes and this interdependency has changed. The results indicate that the Brazil basis now is the dominant aggregator of market information. Port values in the United States have become of lesser importance in information flows in this market. This is significant as the CME, and probably others, are exploring developing a separate contract specification, or instrument, specifically tied to Brazil. The vitality of a separate contract specification ultimately depends on intermarket correlations. If the dynamic interdependency following March 2018 persists, creating a separate pricing location would be attractive.

References

Addinsoft. 2020. *XLSTAT statistical and data analysis solution.* New York, USA: Addinsoft. Available at: https://www.xlstat.com.

Alexandersson, H. 1986. "A Homogeneity Test Applied to Precipitation Data." *Journal of Climatology* 6(6).

Almeida, I., and M. Durisin. 2019. "CME Plans to Start Brazilian Soybean Futures with B3 Exchange." *Bloomberg*. Available at: https://www.bloomberg.com/news/articles/2019-11-13/cme-plans-to-start-brazilian-soybean-futures-with-b3-exchange [Accessed April 20, 2020].

Box, G., and G. Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Bradshaw, G.W., and D. Orden. 1990. "Granger Causality from the Exchange Rate to Agricultural Prices and Export Sales." *Western Journal of Agricultural Economics* 15(1):100–110.

Bullock, D.W., and W.W. Wilson. 2020. "Factors Influencing the Gulf and Pacific Northwest (PNW) Soybean Export Basis: An Exploratory Statistical Analysis." *Journal of Agricultural and Resource Economics*:Forthcoming.

Cottrell, A., and R. Lucchetti. 2019. *GNU Regression, Econometric and Time-series Library (GRETL).* Available at: http://gretl.sourceforge.net/.

Data Transmission Network. 2020. *ProphetX*.

Dhuyvetter, K.C., and T.L. Kastens. 1998. "Forecasting Crop Basis: Practical Alternatives." In 1998 NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management. Chicago, IL.

Gómez, V., and A. Maravall. 2001. "Automatic Modeling Methods for Univariate Series." In D. Peña, G. C. Tiao, and R. S. Tsay, eds. *A Course in Time Series Analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 171–201. Available at: http://doi.wiley.com/10.1002/9781118032978.ch7 [Accessed March 25, 2020].

Granger, C.W.J. 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." *Econometrica* 37(3):424–438.

Haigh, M.S., and H.L. Bryant. 2000. "The Effect of Barge and Ocean Freight Price Volatility in International Grain Markets." *Agricultural Economics* 25(1):41–58.

Hannan, E.J., and B.G. Quinn. 1979. "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society: Series B (Methodological)* 41(2):190–195.

Hannan, E.J., and J. Rissanen. 1982. "Recursive Estimation of Mixed Autoregressive-Moving Average Order." *Biometrika* 69(1):81–94.

Hatchett, R.B., B.W. Brorsen, and K.B. Anderson. 2010. "Optimal Length of Moving Average to Forecast Futures Basis." *Journal of Agricultural and Resource Economics* 35(1):18–33.

Hertsgaard, D.J., W.W. Wilson, and B. Dahl. 2019. "Costs and Risks of Testing and Blending for Essential Amino Acids in Soybeans." *Agribusiness* 35(2):265–280.

Jiang, B., and M. Hayenga. 1997. "Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches." In 1997 NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management. Chicago, IL.

Johansen, S. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59(6):1551–1580.

Johansen, S. 1988. "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control* 12(2):231–254.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin. 1992. "Testing the null hypothesis of stationarity against the alternative of a unit root." *Journal of Econometrics* 54(1–3):159–178.

Lakkakula, P. 2018. "Testing causality among five fertilizer prices." *Applied Economics Letters* 25(9):601–606.

Lakkakula, P., and W.W. Wilson. 2020. "Origin and Export Basis Interdependencies in Soybeans: A Panel Data Analysis." *Journal of Agricultural and Resource Economics*:Forthcoming.

Lara-Chavez, A., and C. Alexander. 2006. "The Effects of Hurricane Katrina on Corn, Wheat, and Soybean Futures Basis." In 2006 NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management. St. Louis, MO.

Lee, Y., and B.W. Brorsen. 2017. "Permanent shocks and forecasting with moving averages." *Applied Economics* 49(12):1213–1225.

Monsell, B., M. Otto, G. Caporello, and V. Gomez. 2013. *X-13ARIMA-SEATS*. United States Department of Commerce, United States Census Bureau.

Onel, G., and B. Karali. 2014. "Relative Performance of Semi-Parametric Nonlinear Models in Forecasting Basis." In 2014 AAEA Annual Meeting. Minneapolis, MN.

Palisade Software. 2018. *@Risk: Advanced Risk Analysis for Microsoft Excel and Project*. Palisade Software, Inc. Available at: www.palisade.com/risk/.

Parcell, J.L. 2000. “The Impact of the LDP on Corn and Soybean Basis in Missouri.” In 2000 NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management. Chicago, IL.

Park, H.S., and C.R. Hurburgh. 2002. “Improving the United States Position in World Soybean Meal Trade.” No. 4, Iowa State University.

Refinitiv. 2020. *Eikon*.

Rennison, J., and G. Meyer. 2018. “US Futures Exchange Weighs Launch of Brazil Soyabean Contracts.” *Financial Times*. Available at: https://www.ft.com/content/38de3e6e-e92c-11e8-a34c-663b3f553b35 [Accessed April 20, 2020].

Rezitis, A.N. 2015. “The Relationship Between Agricultural Commodity Prices, Crude Oil Prices and US Dollar Exchange Rates: A Panel VAR Approach and Causality Analysis.” *International Review of Applied Economics* 29(3):403–434.

Saghaian, S.H. 2010. “The impact of the oil sector on commodity prices: correlation or causation?” *Journal of Agricultural and Applied Economics* 42(3):477–485.

Sanders, D.R., and M.R. Manfredo. 2006. “Forecasting Basis Levels in the Soybean Complex: A Comparison of Time Series Methods.” *Journal of Agricultural and Applied Economics* 38(3):513–523.

Shannon, C.E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27(3):379–423.

Siami-Namini, S., and D. Hudson. 2017. “Volatility Spillover Between Oil Prices, US Dollar Exchange Rates and International Agricultural Commodities Prices.” In Southern Agricultural Economics Association 2017 Annual Meeting. Mobile, Alabama. Available at: http://ageconsearch.umn.edu/record/252845/files/Siami.SAEA2017.pdf.

Skadberg, K., W.W. Wilson, R. Larsen, and B. Dahl. 2015. “Spatial Competition, Arbitrage, and Risk in US Soybeans.” *Journal of Agricultural and Resource Economics* 40(3):442.

Tableau Software. 2020. *Tableau Desktop: Professional Edition*. Seattle, WA: Tableau Software, Inc.

Taylor, M.R., K.C. Dhuyvetter, and T.L. Kastens. 2006. “Forecasting Crop Basis Using Historical Averages Supplemented with Current Market Information.” *Journal of Agricultural and Resource Economics* 31(3):549–567.

Taylor, P.D., and W.G. Tomek. 1984. “Forecasting the Basis for Corn in Western New York.” *Journal of the Northeastern Agricultural Economics Council* 13(1):97–102.

Thakur, M., and C.R. Hurburgh. 2007. “Quality of US Soybean Meal Compared to the Quality of Soybean Meal from Other Origins.” *Journal of the American Oil Chemists’ Society* 84(9):835–843.

Thompson, N.M., A.J. Edwards, J.R. Mintert, and C.A. Hurt. 2019. "Practical Alternatives for Forecasting Corn and Soybean Basis in the Eastern Corn Belt throughout the Crop-Marketing Year." *Journal of Agricultural and Resource Economics* 44(3):571–590.

Tilley, D.S., and S.K. Campbell. 1988. "Performance of the Weekly Gulf-Kansas City Hard-Red Winter Wheat Basis." *American Journal of Agricultural Economics* 70(4):929–935.

Toda, H.Y., and T. Yamamoto. 1995. "Statistical inference in vector autoregressions with possibly integrated processes." *Journal of Econometrics* 66(1–2):225–250.

USDA-AMS. 2019. "Brazil Soybean Transportation." Available at: https://www.ams.usda.gov/sites/default/files/media/Brazil20193rdQtr.pdf [Accessed April 20, 2020].

USDA-OCE, and USDA-AMS. 2015. "Rail Service Challenges in the Upper Midwest: Implications for Agricultural Sectors -- Preliminary Analysis of the 2013-14 Situation." United States Department of Agriculture. Available at: https://www.usda.gov/oce/economics/papers/Rail_Service_Challenges_in_the_Upper_Midwest.pdf.

Ward, J.H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58(301):236–244.

Wilson, W.W., B. Dahl, and D. Hertsgaard. 2020. "Soybean quality differentials, blending, testing and spatial arbitrage." *Journal of Commodity Markets*:Forthcoming.

Wold, H. 1973. "Non-linear Iterative PArtial Least Squares (NIPALS) Modelling: Some Current Developments." In P. R. Krishnaiaah, ed. *Multivariate Analysis III*. New York: Academic Press, pp. 383–407.

Zhang, R., and J. Houston. 2005. "Effects of Price Volatility and Surging South American Soybean Production on Short-Run Basis Dynamics." In 2005 NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management. St. Louis, MO.

Appendix A – Tables

Table 1. Listing of Data Sources for Weekly United States Origin Basis Data

State	City	Platform[a]	Code Used for Data	Notes	Other Codes Used
IA	Alden	DTN	BEANS$.172	Other Code Used for Missing Data	BEANS$.481
IA	Bayard	DTN	BEANS$.292		
IA	Creston	DTN	BEANS$.765		
IA	Dubuque	DTN	BEANS$.884		
IA	Ida Grove	DTN	BEANS$.1453		
IA	Odebolt	DTN	BEANS$.2110		
IA	Red Oak	DTN	BEANS$.4183		
IL	Cairo	DTN	BEANS$.519		
IL	Mound City	DTN	BEANS$.1982		
IL	Nauvoo	DTN	BEANS$.2025		
IN	Aurora	DTN	BEANS$.264		
IN	Jeffersonville	DTN	BEANS$.1500		
IN	Mount Vernon	DTN	BEANS$.3907	Other Code Used for Missing Data	BEANS$.1998
MN	Breckenridge	DTN	BEANS$.426	Other Code Used for Missing Data	BEANS$.425
MN	Jasper	DTN	BEANS$.1492		
ND	Alton	EIK	SOYBALTATN-C1	DTN Used for Missing Data (Other Code)	BEANS$.199
ND	Ayr	DTN	BEANS$.275		
ND	Finley	DTN	BEANS$.1041		
ND	Jamestown	EIK	SOYBGVLJAM-C1		
NE	Beatrice	DTN	BEANS$.301		
NE	Bradshaw	DTN	SOYBPATBDS-C1	DTN Used for Missing Data (Other Code)	BEANS$.3351
NE	Edison	DTN	BEANS$.931		
NE	Fremont	DTN	BEANS$.1093		
NE	Hastings	DTN	BEANS$.3927		
NE	Maywood	DTN	BEANS$.1844		
OH	Cincinnati	DTN	BEANS$.642	Cargill	
OH	Cincinnati	DTN	BEANS$.644	Consolidated Grain and Barge (CGB)	
SD	Madison	DTN	BEANS$.1764		
SD	Marion	DTN	SOYBCFCMRN-C1	DTN Used for Missing Data (Other Code)	BEANS$.1801
SD	Mellette	DTN	BEANS$.1869		
SD	Wolsey	DTN	BEANS$.2946		

[a]DTN = *ProphetX* by Data Transmission Network, EIK = *Eikon* by Refinitiv.

Table 2. Seasonal Analogs by Location Based Upon AHC Clustering Analysis

Location	Analog Assignments by Marketing Year Beginning September 1st[a]															Count
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	
Gulf	1	1	2	1	3	3	1	1	1	***4***	2	3	3	1	1	4
PNW	1	1	1	2	1	1	1	2	1	2	***3***	1	1	1	1	3
Santos	1	1	1	1	1	1	1	***2***	3	3	1	1	1	1	***4***	4
Alden, IA	1	1	1	2	2	1	1	2	***3***	***4***	***5***	1	1	1	1	5
Alton, ND	1	2	1	2	3	2	2	3	***4***	***5***	2	2	2	2	2	5
Aurora, IN	1	1	1	1	1	2	1	3	3	3	2	1	1	1	1	3
Ayr, ND	1	1	1	1	1	1	1	1	***2***	***3***	1	1	1	1	1	3
Bayard, IA	1	1	1	2	2	1	2	2	***3***	***4***	1	1	1	2	1	4
Beatrice, NE	1	1	***2***	3	4	4	4	3	***5***	***6***	4	***7***	1	4	1	7
Bradshaw, NE	1	1	1	2	2	1	1	2	***3***	***4***	1	1	1	1	1	4
Breckenridge, MN	1	1	1	2	1	3	2	2	***4***	***5***	3	1	3	3	3	5
Cairo, IL	1	1	1	2	2	2	2	3	3	3	2	2	2	2	2	3
Cin (Cargill)	1	1	1	2	2	3	2	2	4	4	3	2	1	2	3	4
Cin (CBH)	1	1	1	1	1	2	1	3	3	3	2	1	1	1	1	3
Creston, IA	1	1	1	1	1	1	1	1	2	2	***3***	1	1	1	1	3
Dubuque, IA	1	1	1	2	***3***	4	1	***5***	***6***	***7***	***8***	2	4	4	4	8
Edison, NE	1	1	1	2	3	1	2	2	***4***	***5***	1	1	1	1	1	5
Finley, ND	1	2	1	3	1	2	2	3	***4***	***5***	2	2	2	2	2	5
Freemont, NE	1	1	1	2	1	1	1	2	3	3	1	1	1	1	1	3
Hastings, NE	1	1	1	***2***	3	4	3	3	***5***	***6***	4	1	4	1	4	6
Ida Grove, IA	1	1	1	1	1	1	1	***2***	3	3	1	1	1	1	1	3
Jamestown, ND	1	1	***2***	3	3	1	1	3	***4***	***5***	1	1	1	1	1	5
Jasper, MN	1	1	1	2	2	2	2	***3***	***4***	***5***	***6***	1	2	2	2	6
Jeffersonville, IN	1	1	1	1	2	2	2	3	***4***	3	2	2	2	2	2	4
Madison, SD	1	1	1	2	1	1	2	2	***3***	***4***	2	1	1	1	1	4
Marion, SD	1	1	1	1	1	1	1	1	2	2	***3***	1	1	1	1	3
Maywood, NE	1	1	1	1	***2***	1	1	1	3	3	1	1	1	1	1	3
Mellette, SD	1	1	1	1	1	1	1	1	***2***	***3***	1	1	1	1	1	3
Mound City, IL	1	1	***2***	3	4	3	3	4	***5***	***6***	3	3	3	3	3	6
Mount Vernon, IN	1	1	1	1	1	1	1	2	***3***	2	1	1	1	1	1	3
Nauvoo, IL	1	1	1	2	2	1	1	2	***3***	***4***	1	2	1	1	1	4
Odebolt, IA	1	1	1	2	1	1	1	2	3	3	1	1	1	1	1	3
Red Oak, IA	1	1	1	2	2	3	2	2	***4***	***5***	***6***	3	3	3	3	6
Wolsey, SD	1	2	1	1	2	2	2	2	***3***	***4***	2	2	2	2	2	4

[a]Values in bold italics indicate single year analogs.

Table 3. Result of Co-occurrence Clustering of Locations by Marketing Year and Seasonal Analog Identifier

Class	**1**	**2**	**3**	**4**
Objects	21	9	2	2
Within-class variance	5.29	6.47	14.50	18.00
Locations in Class	Gulf	Alton, ND	Beatrice, NE	Dubuque, IA
	PNW	Breckenridge, MN	Mound City, IL	Hastings, NE
	Santos	Cairo, IL		
	Alden, IA	Cin (Cargill)		
	Aurora, IN	Finley, ND		
	Ayr, ND	Jasper, MN		
	Bayard, IA	Jeffersonville, IN		
	Bradshaw, NE	Red Oak, IA		
	Cin (CBH)	Wolsey, SD		
	Creston, IA			
	Edison, NE			
	Freemont, NE			
	Ida Grove, IA			
	Jamestown, ND			
	Madison, SD			
	Marion, SD			
	Maywood, NE			
	Mellette, SD			
	Mount Vernon, IN			
	Nauvoo, IL			
	Odebolt, IA			

Table 4. Result of Co-occurrence Clustering of Marketing Years by Locations and Seasonal Analog Identifier

Class	**1**	**2**	**3**
Objects	13	1	1
Within-class variance	23.85	0.00	0.00
Marketing Years in Class	2004/05	2012/13	2013/14
	2005/06		
	2006/07		
	2007/08		
	2008/09		
	2009/10		
	2010/11		
	2011/12		
	2014/15		
	2015/16		
	2016/17		
	2017/18		
	2018/19		

Table 5. Results of Standard Normal Homogeneity Test (SNHT) For Each Location (Sep 2008 to Dec 2019)

Location[a]	T_0	Date	p-value	Mu1		Mu2		Change	
Gulf	**34.38**	**Sep-16**	**0.0500**	**$**	**0.8060**	**$**	**0.3980**	**$**	**(0.4080)**
PNW	**40.29**	**Dec-14**	**< 0.0001**	**$**	**1.2320**	**$**	**0.8640**	**$**	**(0.3680)**
Santos, Brazil	**24.61**	**Aug-14**	**0.0026**	**$**	**0.2240**	**$**	**0.9320**	**$**	**0.7080**
Alden, IA	**27.16**	**Sep-15**	**0.0190**	**$**	**(0.2670)**	**$**	**(0.7240)**	**$**	**(0.4570)**
Alton, ND	**33.79**	**Sep-16**	**0.0160**	**$**	**(0.6700)**	**$**	**(1.0720)**	**$**	**(0.4020)**
Aurora, IN	11.59	Nov-11	0.1083	$	(0.1506)	$	0.1801	$	0.3307
Ayr, ND	**30.02**	**Aug-14**	**0.0173**	**$**	**(0.6180)**	**$**	**(0.9610)**	**$**	**(0.3430)**
Bayard, IA	**26.57**	**Sep-15**	**0.0196**	**$**	**(0.2610)**	**$**	**(0.7070)**	**$**	**(0.4460)**
Beatrice, NE	**31.07**	**Aug-15**	**0.0179**	**$**	**(0.4090)**	**$**	**(0.8390)**	**$**	**(0.4300)**
Bradshaw, NE	**41.03**	**Sep-14**	**0.0004**	**$**	**(0.3490)**	**$**	**(0.8380)**	**$**	**(0.4890)**
Breckenridge, MN	**36.65**	**Feb-16**	**0.0004**	**$**	**(0.5440)**	**$**	**(0.9820)**	**$**	**(0.4380)**
Cairo, IL	**23.09**	**Oct-10**	**0.0190**	**$**	**(0.1450)**	**$**	**0.3710**	**$**	**0.5160**
Cin CGB, OH	11.94	Nov-11	0.1026	$	(0.1446)	$	0.1929	$	0.3375
Cin Cargill, OH	12.95	Nov-11	0.0701	$	(0.1488)	$	0.1955	$	0.3443
Creston, IA	**36.06**	**Sep-15**	**0.0005**	**$**	**(0.1370)**	**$**	**(0.7150)**	**$**	**(0.5780)**
Dubuque, IA	**40.41**	**Sep-16**	**0.0001**	**$**	**(0.1400)**	**$**	**(0.7860)**	**$**	**(0.6460)**
Edison, NE	**45.31**	**Feb-16**	**0.0001**	**$**	**(0.4650)**	**$**	**(1.0400)**	**$**	**(0.5750)**
Finley, ND	**41.04**	**Jan-14**	**0.0002**	**$**	**(0.6270)**	**$**	**(1.0050)**	**$**	**(0.3780)**
Fremont, NE	**24.04**	**Sep-15**	**0.0204**	**$**	**(0.1170)**	**$**	**(0.5350)**	**$**	**(0.4180)**
Hastings, NE	**37.97**	**Sep-15**	**0.0170**	**$**	**(0.3350)**	**$**	**(0.8700)**	**$**	**(0.5350)**
Ida Grove, IA	**26.67**	**Sep-15**	**0.0201**	**$**	**(0.2640)**	**$**	**(0.6920)**	**$**	**(0.4280)**
Jamestown, ND	**46.95**	**Nov-13**	**< 0.0001**	**$**	**(0.5960)**	**$**	**(1.0140)**	**$**	**(0.4180)**
Jasper, MN	**32.97**	**Aug-15**	**0.0017**	**$**	**(0.3370)**	**$**	**(0.8330)**	**$**	**(0.4960)**
Jeffersonville, IN	11.72	Nov-11	0.1060	$	(0.1642)	$	0.1570	$	0.3213
Madison, SD	**31.00**	**Aug-14**	**0.0175**	**$**	**(0.4330)**	**$**	**(0.8270)**	**$**	**(0.3940)**
Marion, SD	**31.83**	**Sep-15**	**0.0174**	**$**	**(0.4120)**	**$**	**(0.8850)**	**$**	**(0.4730)**
Maywood, NE	**47.59**	**Sep-14**	**< 0.0001**	**$**	**(0.4720)**	**$**	**(1.0650)**	**$**	**(0.5930)**
Mellette, SD	**36.48**	**Sep-14**	**0.0008**	**$**	**(0.5190)**	**$**	**(0.9490)**	**$**	**(0.4300)**
Mound City, IL	**18.17**	**Nov-11**	**0.0273**	**$**	**(0.1150)**	**$**	**0.2420**	**$**	**0.3570**
Mount Vernon, IN	**15.16**	**Sep-16**	**0.0439**	**$**	**0.2820**	**$**	**(0.0900)**	**$**	**(0.3720)**
Nauvoo, IL	12.75	Sep-16	0.0911	$	0.0148	$	(0.3073)	$	(0.3221)
Odebolt, IA	**27.93**	**Sep-15**	**0.0191**	**$**	**(0.2780)**	**$**	**(0.7250)**	**$**	**(0.4470)**
Red Oak, IA	**36.58**	**Dec-15**	**0.0007**	**$**	**(0.1330)**	**$**	**(0.7130)**	**$**	**(0.5800)**
Wolsey, SD	**40.68**	**Sep-14**	**0.0004**	**$**	**(0.5080)**	**$**	**(0.9680)**	**$**	**(0.4600)**

[a]Values in bold have p-values of 0.05 or less (i.e., 95% significance).

Table 6. Summary of Box-Jenkins Model Identification by Location with Test Results for ARCH and China Trade Dispute Intervention Dummy

Location	Box-Jenkins Model Identification[a]	Adjusted R^2	ARCH Test (12 Lags)[b] LM	ARCH Test (12 Lags)[b] p-value	China Dispute Dummy D1 (3/2018 to 12/2019) Coeff	China Dispute Dummy D1 (3/2018 to 12/2019) p-value
Gulf	SARIMA (2,0,2)x(1,0,0)	0.671	13.12	0.3607	$ (0.0441)	0.7777
PNW	ARIMA(1,1,1)	0.765	19.43	0.0787	$ 0.0436	0.7990
Santos, Brazil	SARIMA (3,0,1)x(0,1,1)	0.696	13.09	0.3626	**$ 0.6389**	**0.0261**
Alden, IA	SARIMA (1,0,1)x(0,1,1)	0.725	**56.62**	**<0.0001**	$ (0.0485)	0.8031
Alton, ND	SARIMA (2,0,2)x(0,0,1)	0.631	**80.54**	**<0.0001**	$ (0.1473)	0.3747
Aurora, IN	SARIMA (1,0,1)x(1,0,0)	0.729	**62.82**	**<0.0001**	$ (0.0686)	0.7353
Ayr, ND	SARIMA (2,0,2)x(1,0,0)	0.659	**77.85**	**<0.0001**	$ (0.0730)	0.6537
Bayard, IA	SARIMA (1,0,0)x(1,0,0)	0.798	**77.27**	**<0.0001**	$ (0.0742)	0.6613
Beatrice, NE	SARIMA (1,0,0)x(1,0,0)	0.734	**81.23**	**<0.0001**	$ (0.0728)	0.6674
Bradshaw, NE	SARIMA (1,0,0)x(1,0,0)	0.750	**62.21**	**<0.0001**	$ (0.1425)	0.3880
Breckenridge, MN	SARIMA (2,0,2)x(1,0,0)	0.718	**64.60**	**<0.0001**	$ (0.1772)	0.2530
Cairo, IL	SARIMA (1,0,0)x(1,0,0)	0.745	**73.24**	**<0.0001**	$ (0.0422)	0.8219
Cin Cargill, OH	SARIMA (3,0,1)x(1,0,0)	0.687	**56.74**	**<0.0001**	$ (0.0382)	0.8322
Cin CGB, OH	SARIMA (3,0,1)x(1,0,1)	0.701	**46.13**	**<0.0001**	$ (0.0726)	0.7490
Creston, IA	SARIMA (3,0,1)x(1,0,1)	0.777	**56.67**	**<0.0001**	$ (0.1445)	0.4088
Dubuque, IA	SARIMA (1,0,0)x(1,0,0)	0.731	**42.64**	**<0.0001**	$ (0.1822)	0.3599
Edison, NE	SARIMA (2,0,2)x(1,0,0)	0.713	**81.97**	**<0.0001**	$ (0.2620)	0.1051
Finley, ND	ARIMA(1,0,0)	0.581	**69.40**	**<0.0001**	$ (0.2480)	0.0876
Fremont, NE	SARIMA (1,0,1)x(1,0,0)	0.723	**72.86**	**<0.0001**	$ (0.1505)	0.3688
Hastings, NE	SARIMA (1,0,0)x(1,0,0)	0.701	**72.21**	**<0.0001**	$ (0.2002)	0.2655
Ida Grove, IA	SARIMA (1,0,1)x(0,0,1)	0.826	**68.09**	**<0.0001**	$ (0.1279)	0.4125
Jamestown, ND	SARIMA (2,0,2)x(0,0,1)	0.666	**94.24**	**<0.0001**	$ (0.2064)	0.1624
Jasper, MN	SARIMA (1,0,0)x(1,0,0)	0.734	**37.88**	**0.0002**	$ (0.1382)	0.4688
Jeffersonville, IN	SARIMA (1,0,0)x(1,0,0)	0.719	**73.07**	**<0.0001**	$ (0.0739)	0.6973
Madison, SD	SARIMA (2,0,2)x(1,0,0)	0.715	**55.48**	**<0.0001**	$ (0.1549)	0.2627
Marion, SD	SARIMA (1,0,0)x(1,0,0)	0.736	**37.00**	**0.0002**	$ (0.1654)	0.3580
Maywood, NE	SARIMA (1,0,0)x(1,0,0)	0.745	**79.44**	**<0.0001**	$ (0.1659)	0.3956
Mellette, SD	SARIMA (3,0,1)x(1,0,1)	0.790	**68.37**	**<0.0001**	$ 0.0690	0.6626
Mound City, IL	SARIMA (1,0,0)x(1,0,0)	0.685	**83.99**	**<0.0001**	$ (0.0683)	0.6918
Mount Vernon, IN	SARIMA (1,0,0)x(1,0,0)	0.758	**69.26**	**<0.0001**	$ (0.0912)	0.6691
Nauvoo, IL	SARIMA (1,0,1)x(1,0,0)	0.776	**82.24**	**<0.0001**	$ (0.2062)	0.2305
Odebolt, IA	SARIMA (1,0,1)x(1,0,0)	0.798	**70.37**	**<0.0001**	$ (0.1200)	0.4702
Red Oak, IA	SARIMA (1,0,0)x(1,0,0)	0.749	**32.23**	**0.0013**	$ (0.1686)	0.3844
Wolsey, SD	SARIMA (2,0,2)x(0,0,1)	0.769	**62.96**	**<0.0001**	$ 0.0160	0.9190

[a]Based upon application of SEATS procedure (Gómez and Maravall, 2001).

[b]H0: No ARCH effect is present.

[c]Dummy variable incorporated into ARIMA/SARIMA model. D equals 1 over entire time period indicated.

Table 7. Toda-Yamamoto (1995) Asymptotic Wald Statistics for Pre-China Trade War Period

	Dependent[a]				
Independent	Gulf	PNW	Santos, Brazil	Eastern Corn Belt	Western Corn Belt
Gulf	#N/A	12.28***	11.92***	5.87	14.53***
PNW	13.39***	#N/A	4.49	0.44	2.77
Santos, Brazil	13.13***	13.91***	#N/A	9.05**	15.52***
Eastern Corn Belt	72.91***	10.80**	15.57***	#N/A	7.82**
Western Corn Belt	78.87***	2.32	15.75***	21.74***	#N/A

[a]Distributed as a chi-squared distribution with 3 degrees of freedom. Weekly series prior to March 18, 2018.

Table 8. Toda-Yamamoto (1995) Asymptotic Wald Statistics for Post-China Trade War Period

	Dependent[a]				
Independent	Gulf	PNW	Santos, Brazil	Eastern Corn Belt	Western Corn Belt
Gulf	#N/A	4.00	0.33	4.82	1.12
PNW	4.86	#N/A	0.49	0.24	1.16
Santos, Brazil	1.16	1.32	#N/A	1.27	1.14
Eastern Corn Belt	0.99	0.22	19.53***	#N/A	0.48
Western Corn Belt	1.12	1.31	44.47***	0.89	#N/A

[a]Distributed as a chi-squared distribution with 3 degrees of freedom. Weekly series following March 18, 2018.

Appendix B – Figures

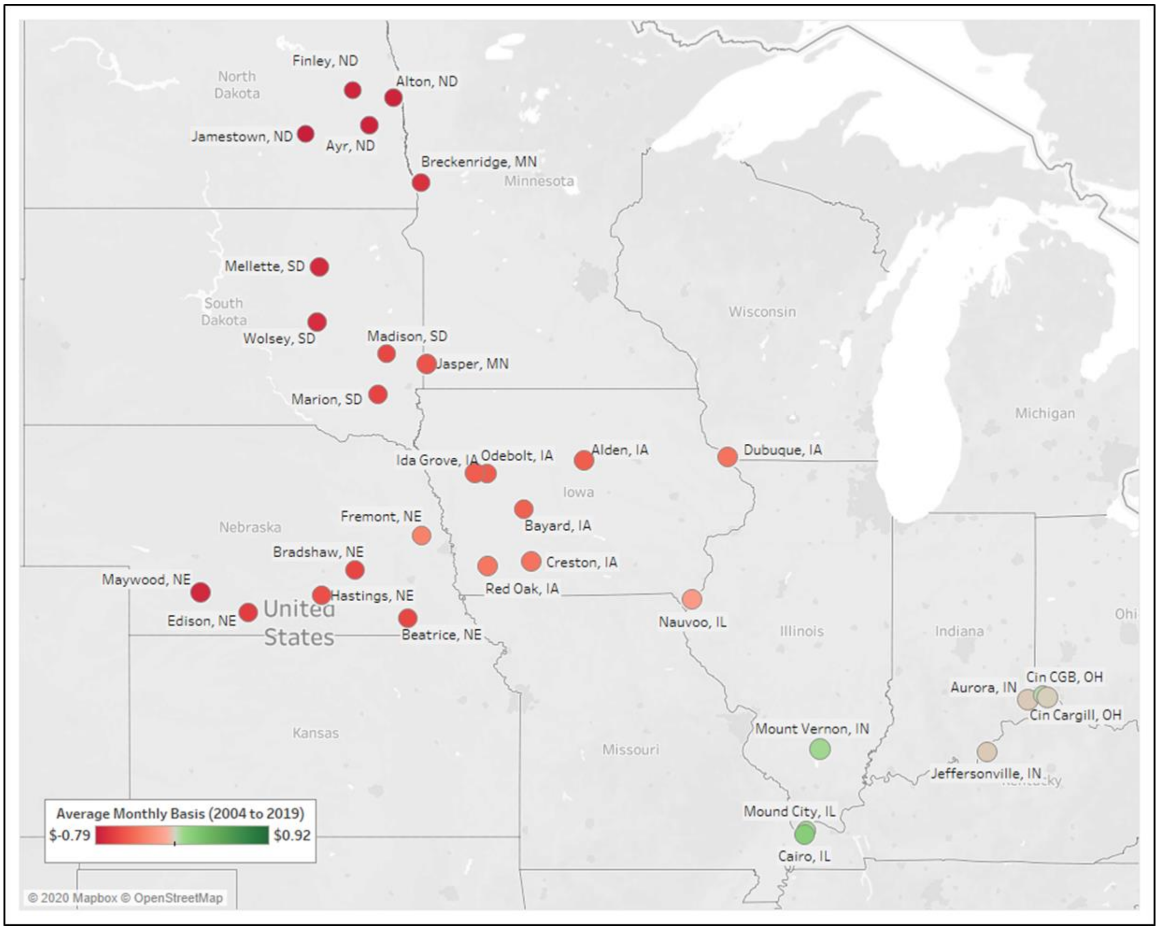

Figure 1. Interior locations used in study (with average monthly basis from 2004 to 2019).

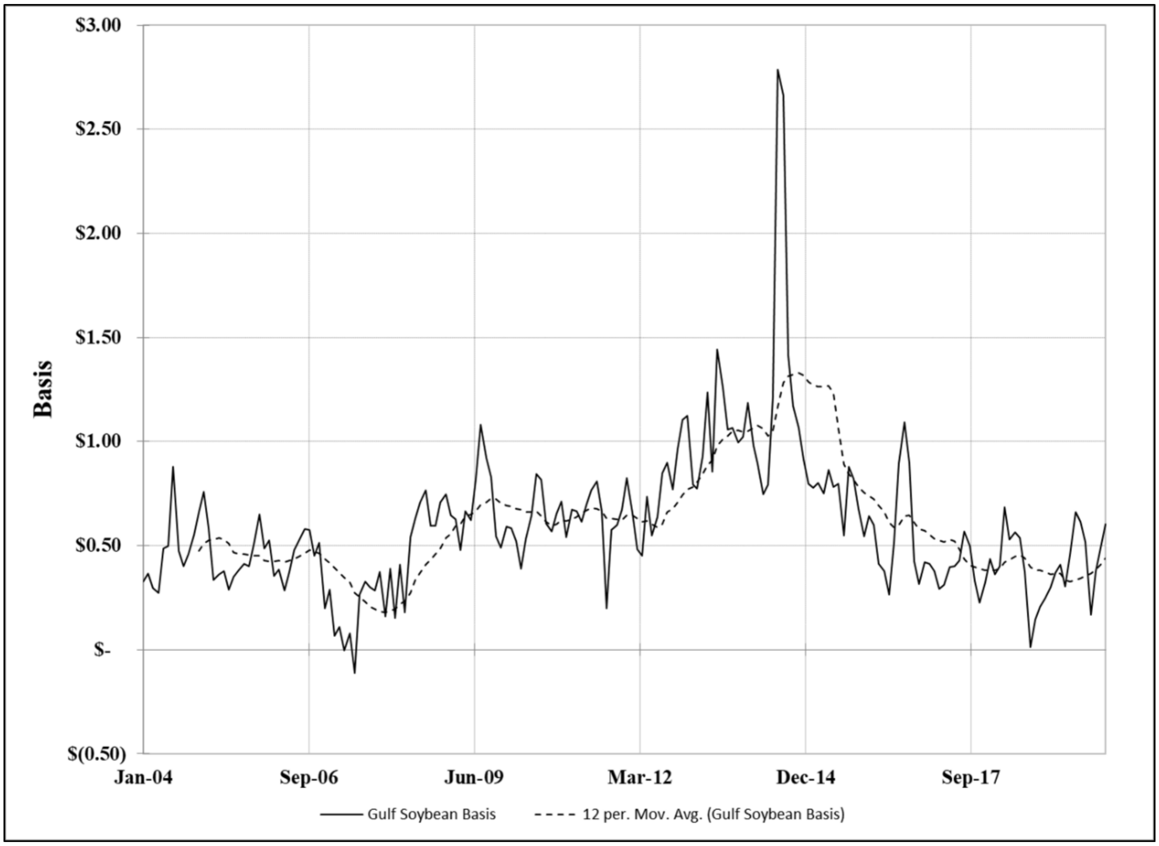

Figure 2. Monthly average USG soybean basis (Jan 2004 to Dec2019) with 12-month moving average.

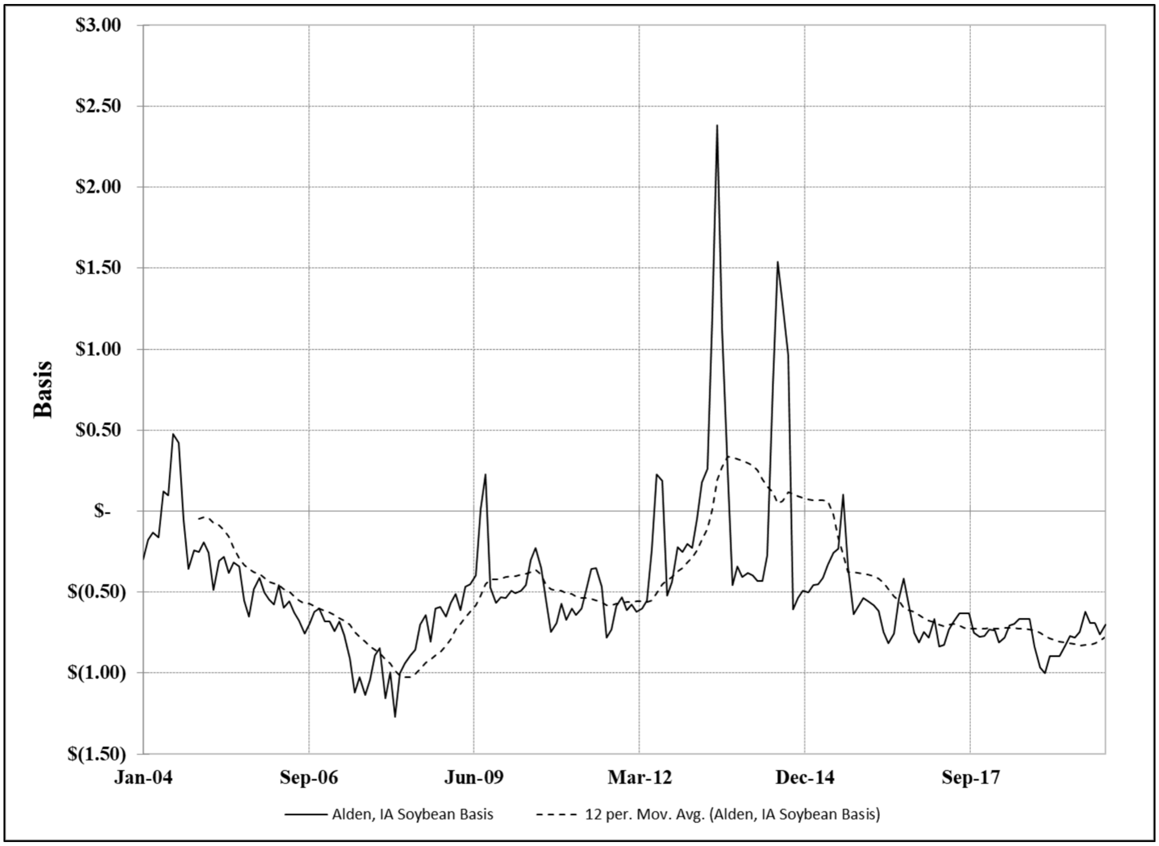

Figure 3. Monthly average Alden, IA soybean basis (Jan 2004 to Dec 2019) with 12-month moving average.

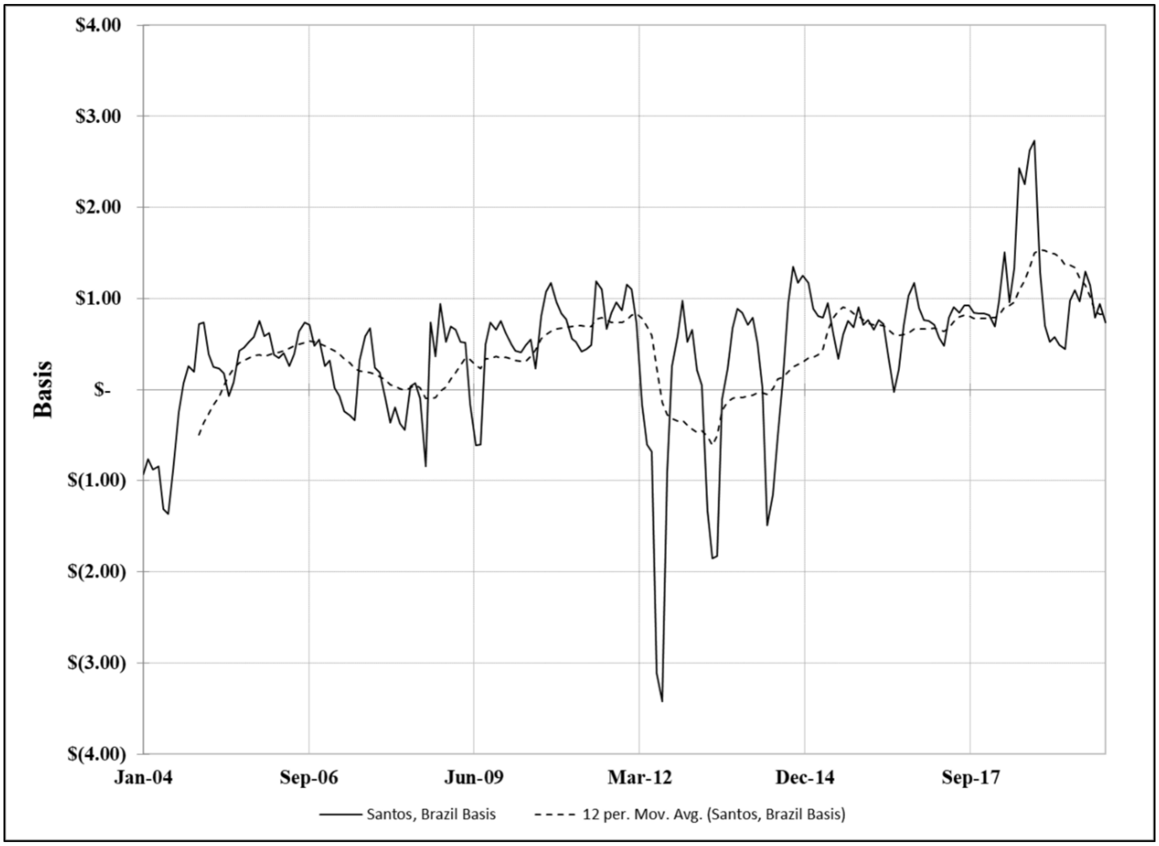

Figure 4. Monthly average Santos, Brazil soybean basis (Jan 2004 to Dec 2019) with 12-month moving average.

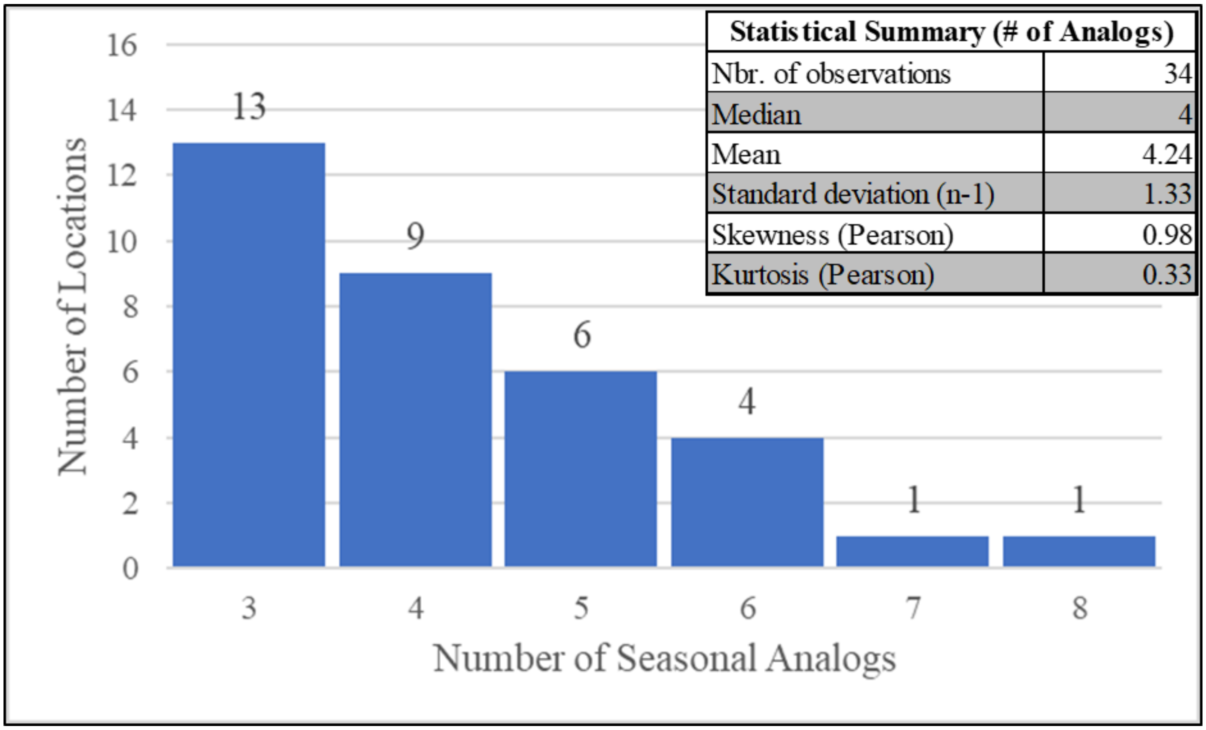

Figure 5. Histogram of Number of Seasonal Analogs by Location with Statistical Summary

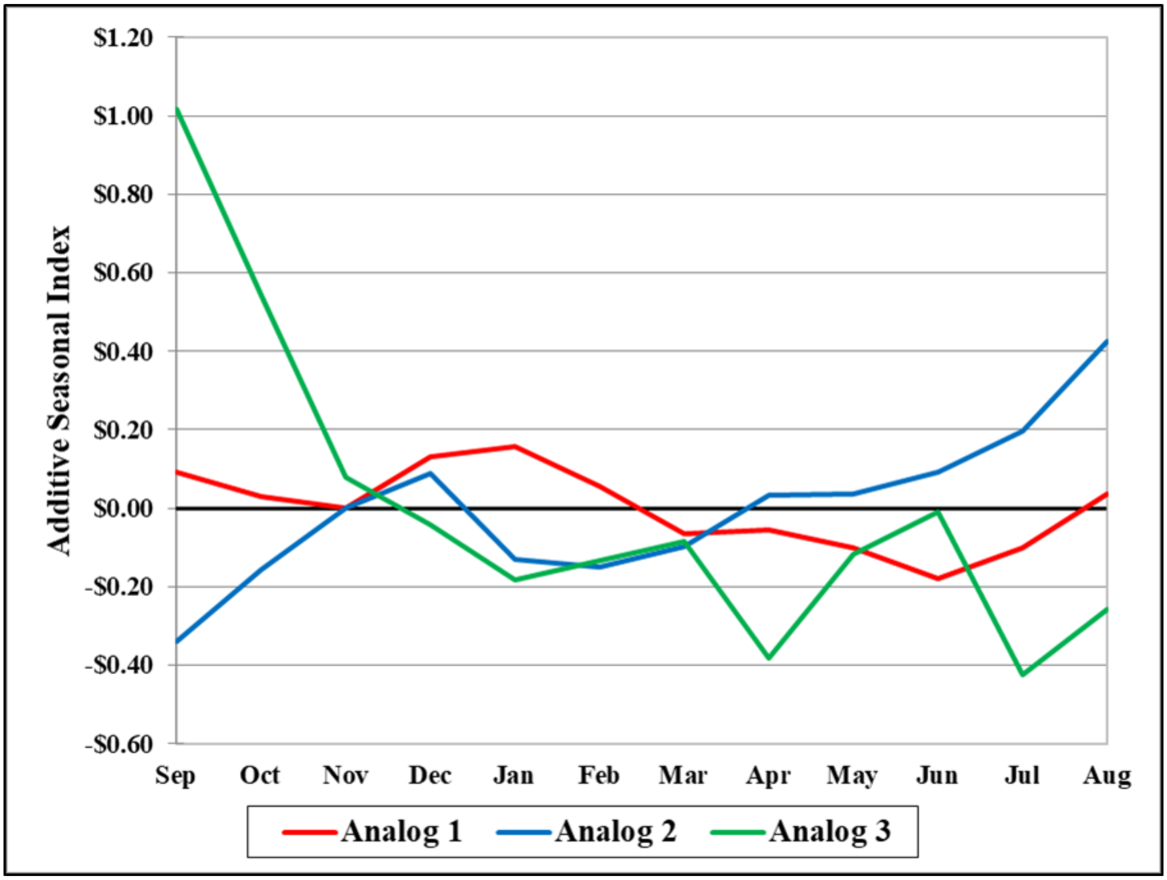

Figure 6. Plot of Seasonal Analogs for PNW Soybean Basis

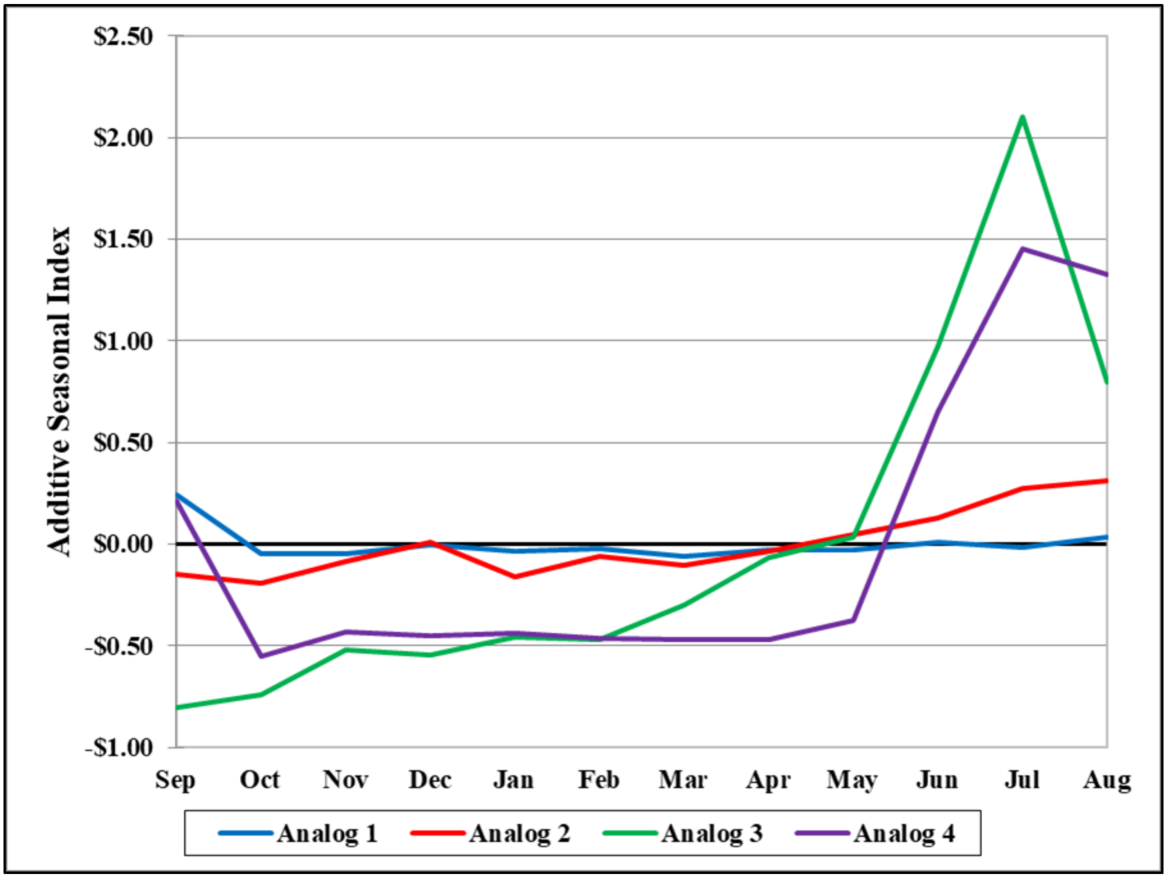

Figure 7. Plot of Seasonal Analogs for Bayard, Iowa Soybean Basis

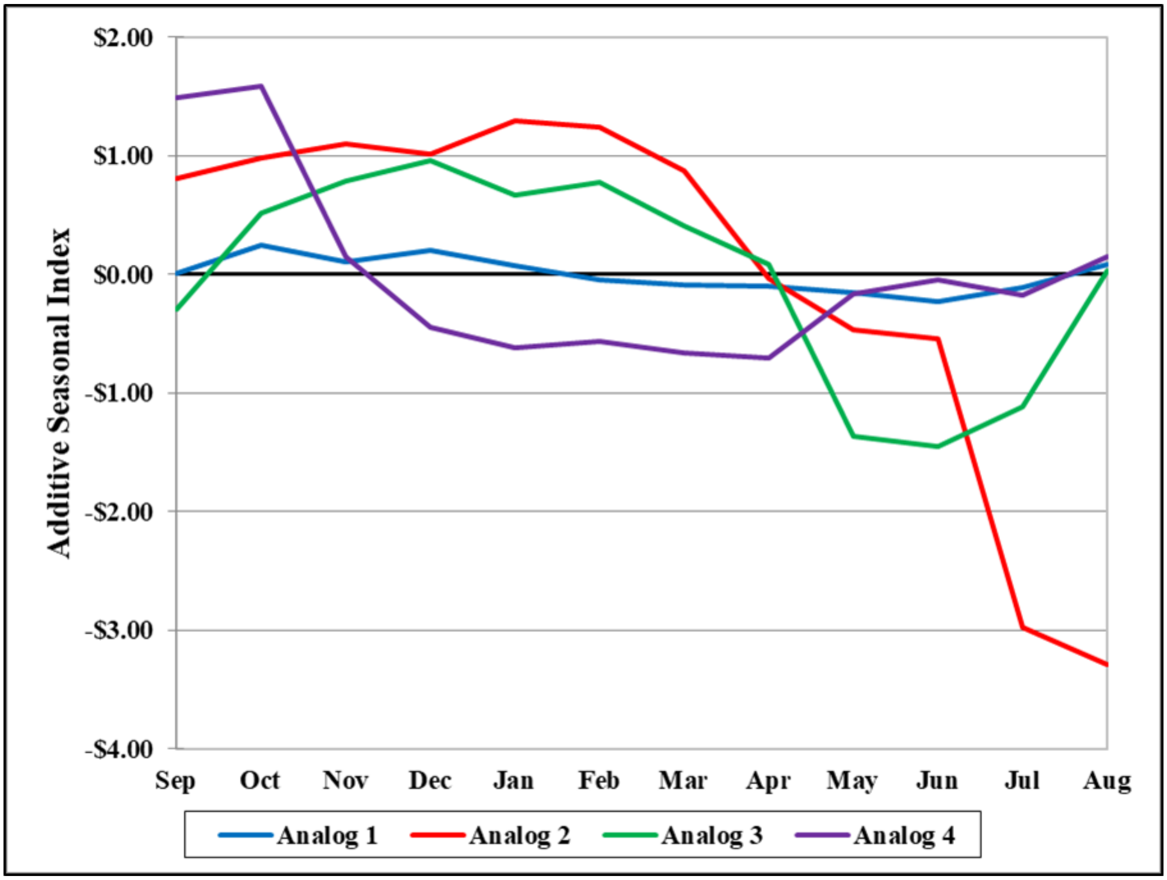

Figure 8. Plot of Seasonal Analogs for Santos, Brazil Soybean Basis

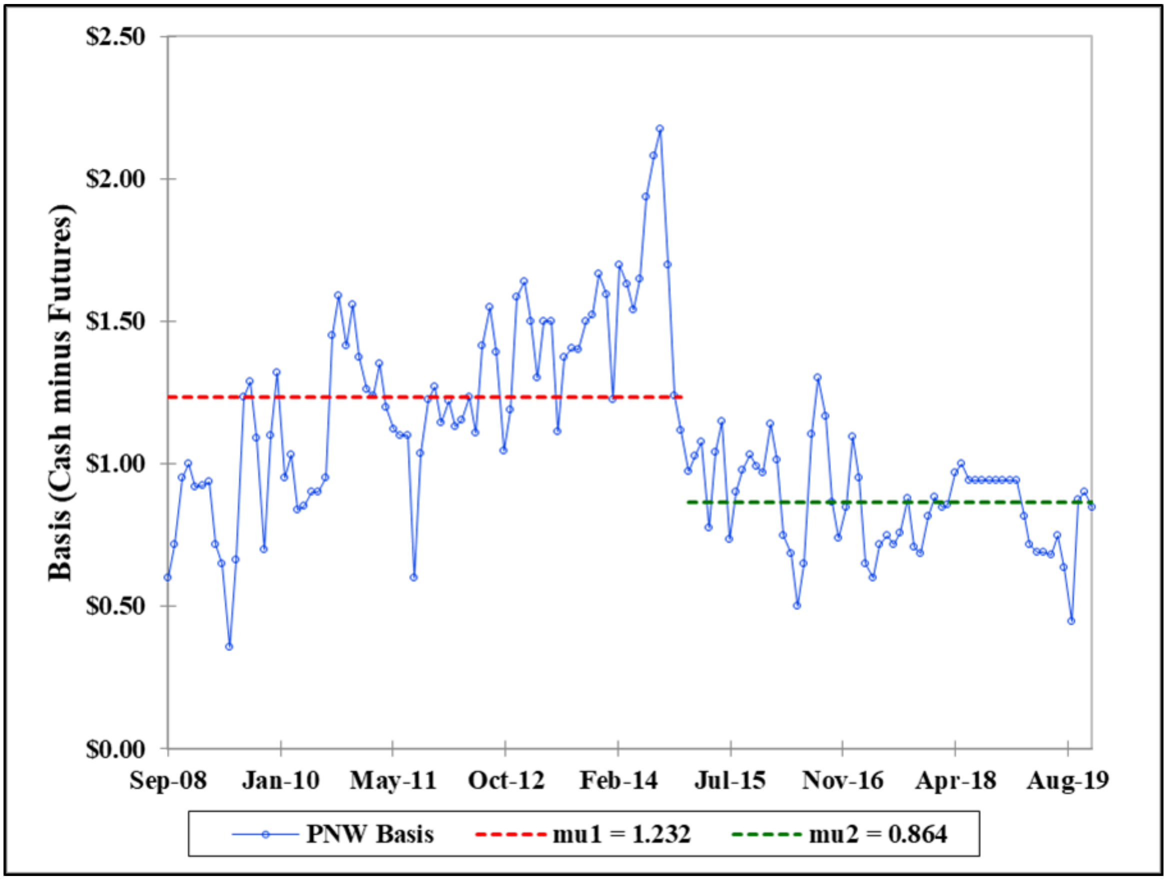

Figure 9. SNHT Breakpoint for PNW Soybean Basis

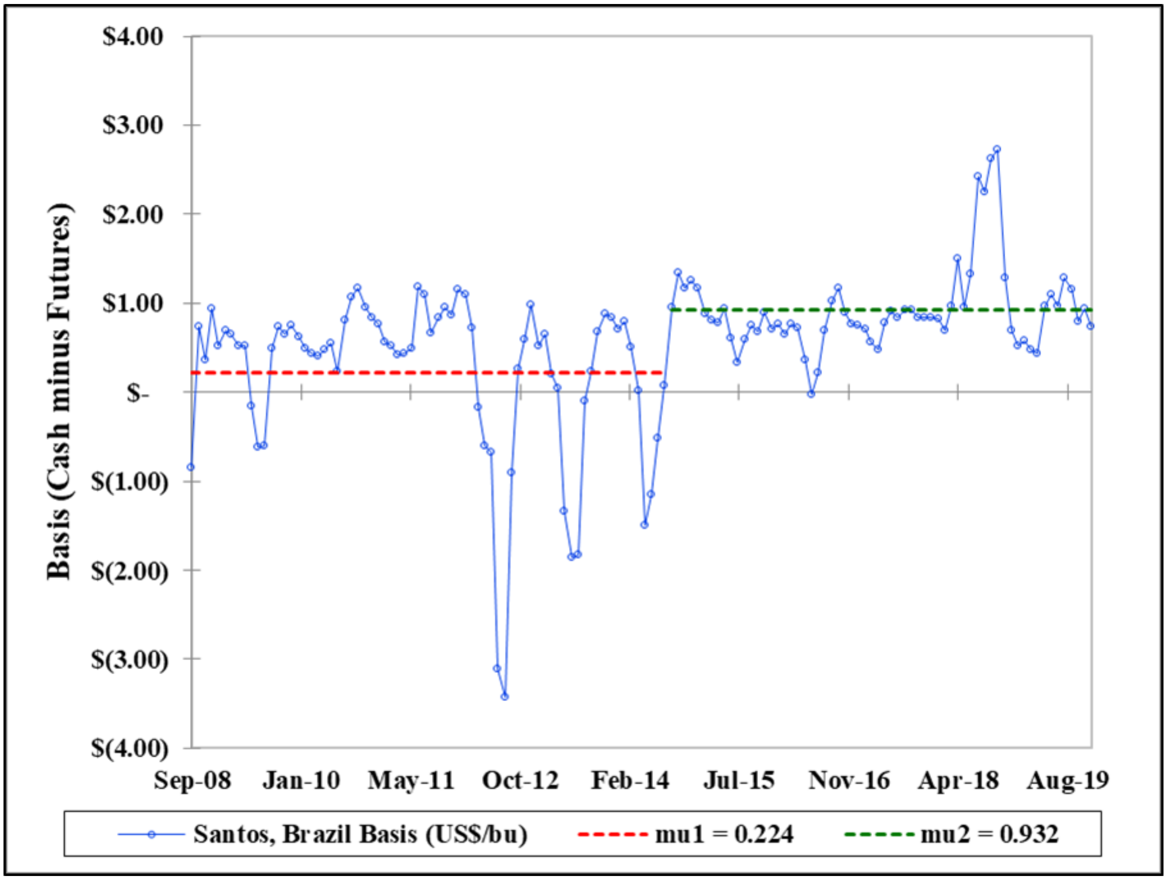

Figure 10. SNHT Breakpoint for Santos, Brazil Soybean Basis

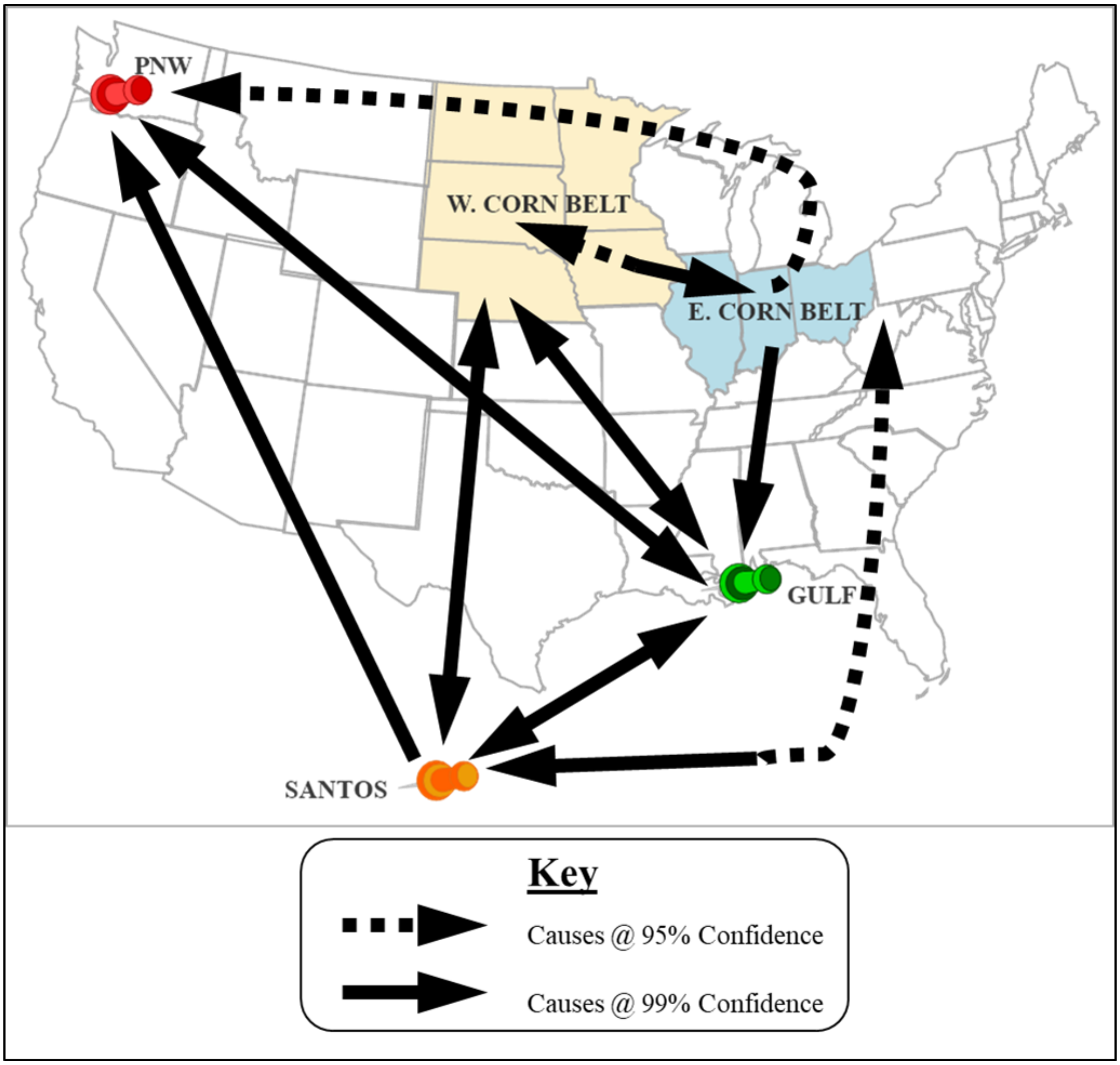

Figure 11. Map of Causal Flows for Pre-China Trade War Time Period
(Prior to March 18, 2018)

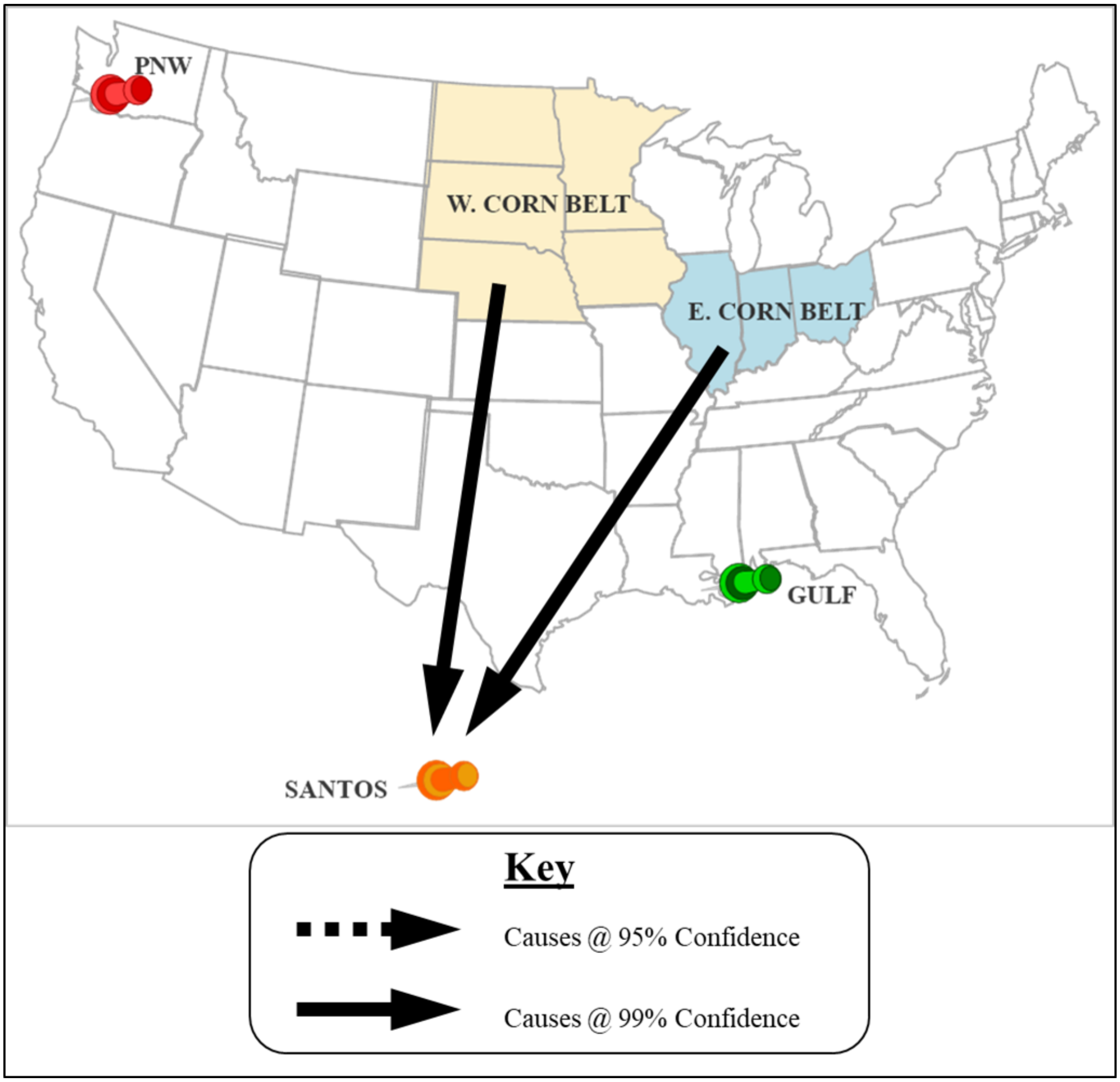

Figure 12. Map of Causal Flows for Post-China Trade War Time Period (Following March 18, 2018)