

## **Factors Influencing the Gulf and Pacific Northwest (PNW) Soybean Export Basis: An Exploratory Statistical Analysis**

by

David W. Bullock and William W. Wilson

Suggested citation format:

Bullock, D. W. and W. W. Wilson. 2019. "Factors Influencing the Gulf and Pacific Northwest (PNW) Soybean Export Basis: An Exploratory Statistical Analysis." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Minneapolis, MN. [<http://www.farmdoc.illinois.edu/nccc134>].

**Factors Influencing the Gulf and Pacific Northwest (PNW) Soybean Export Basis:  
An Exploratory Statistical Analysis**

**David W. Bullock and William W. Wilson<sup>1</sup>**

*Paper presented at the NCCC-134 Conference on Applied Commodity Analysis,  
Forecasting, and Market Risk Management  
Minneapolis, Minnesota, April 15-16, 2019*

*Copyright 2019 by David W. Bullock and William W. Wilson. All rights reserved. Readers may  
make verbatim copies of this document for non-commercial purposes by any means, provided  
that this copyright notice appears on all such copies.*

---

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

## **Factors Influencing the Gulf and Pacific Northwest (PNW) Soybean Export Basis: An Exploratory Statistical Analysis**

*Growth in the export marketing of soybeans has drawn attention to the basis volatility in these market channels. Indeed, there has been greater growth in soybean exports compared to other commodities and this is due in part to the growth of exports to China. Concurrently, there has been substantial volatility in the basis at the primary U.S. export locations: the U.S. Gulf and the Pacific Northwest (PNW). The purpose of this study is to examine the impact of supply/demand, export competition and logistical variables on both the average level and seasonality (analog year) of U.S. export basis values for the 2004/05 through 2015/16 marketing years (September through August for U.S. soybeans). The results indicate that the average market year level of the basis is primarily influenced by export competition from Brazil and export demand – particularly from China; however, domestic demand (soybean crush) also has some influence. Rail transportation costs to both the Gulf and PNW have an influence on the basis level; however, barge and ocean freight rates appear to not have a significant influence on the level of the basis. Application of agglomerative hierarchical clustering resulted in the identification of 5 and 4 distinct analogs (over the 12 marketing years in the dataset) for the Gulf and PNW respectively. Application of the two-sample mean difference tests (Lebart, Morineau and Piron 2000) to the analogs indicate that the seasonal analog of the export basis is more heavily influenced by internal logistical conditions (late railcar placement and secondary railcar values), pace of farmer marketings, transportation cost differentials (between ports), and individual port export activity (ships in port and export inspections) rather than international and domestic demand.*

**Key words:** soybean export basis, seasonal analog, partial least squares (PLS), agglomerative hierarchical clustering (AHC), Lebart mean difference test

## **Introduction**

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. This may include allocating commodities into and out of storage, as well as into the marketing system in addition to spatial allocation across destinations. This coordination insures the efficient and orderly marketing of a commodity. These aspects of demand (time, place, and form) are reflected in the components of the basis for a storable commodity.

For participants in the exchange traded commodity markets, the ability to forecast basis is essential for making basic marketing and management decisions (Dhuyvetter and Kastens 1998). Combining basis forecasts with current futures prices is a common method of deriving local cash price forecasts. Additionally, hedging as a price risk management strategy is essentially an exchange of cash price risk for basis risk (Working 1953). Therefore, an understanding of the basis is essential in evaluating potential hedging opportunities. An important feature of the basis

is that it is seasonal, but, the characterization of the seasonal behavior (timing of min, max, volatility, etc.) varies from year to year depending on market fundamentals.

The purpose of this study is to examine the impact of fundamental factors upon both the level (marketing year average) and seasonality (by marketing year) of Gulf and PNW nearby soybean basis values for the 2004/05 through 2015/16 marketing years. Explanatory variables include Brazilian basis values (FOB Paranagua), nearby futures spreads, rail transportation costs (tariff plus fuel surcharge), secondary railcar values (DCV), barge rates, ocean freight rates, number of ships in port (Gulf and PNW), and a number of additional supply / demand variables (too many to list here).

Since the number of explanatory variables (27) exceeds the number of observations (12 marketing years), this study utilizes an exploratory regression technique called partial least squares (PLS) (Herman Wold 1966) to determine the influence of the explanatory variables upon the marketing year average basis level. PLS is similar to principal components regression (PCR); however, it has the advantage of considering information in both the dependent and explanatory variables in constructing its regression components (latent variables). A useful output of the PLS procedure is an index called the Variable Importance in Projection (VIP) where values greater than one are considered important in making the projection (Wold et al. 1993, pp. 523-550). The VIP values can effectively be used to pare down the number of explanatory variables that are fitted in the final round of PLS estimation. Regression coefficient t-statistics are used to test the statistical significance of each explanatory variable in the final PLS estimation.

To explain basis seasonality, additive seasonal indices for each marketing year are calculated by taking the difference between the monthly value and the marketing year average. The marketing years are then grouped into seasonal analogues by applying agglomerative hierarchical clustering (AHC) using Euclidian distance as the clustering metric (Ward 1963). The optimal number of clusters is determined for each location (Gulf and PNW) using minimum entropy. To determine which explanatory variables are most significant to each analog cluster, a two-sample z-test (Lebart, Morineau and Piron 2000) for testing the differences in means between a subset and a larger set containing the subset was used with each analog comprising the subset as the test was applied to each explanatory variable.

In the following section, we provide a review of the previous related studies. This is followed by a sections describing the dataset and methodology, the results of the exploratory statistical analysis, and a summary of the conclusions and implications from this research.

## **Previous Studies on Basis in the Grain and Oilseed Sector**

There are several strains of literature related to this study.<sup>2</sup> One is the empirical analysis of basis. There have been numerous studies on basis behavior, and factors influencing its behavior.

---

<sup>2</sup> This review is not completely exhaustive but provides the major thrust of each of the themes as it pertains to this study.

Second is the impacts of rail (transport) on basis values. While historically these have not been so common, there were a number of studies on this topic concurrent with and following the 2013/14 marketing year. We also provide a brief description of what we refer as “analogue year” analysis that is a concept we promote in the empirical section. However, common in industry empirical analysis, this approach has been less common in the academic literature.

***Basis at the futures delivery market:***

As a precursor to some of the empirical analyses of the basis, it is important that the basis have a highly predetermined behavior when evaluated at the delivery market of the futures contracts. This is fundamental to the studies that follow. The basis at the delivery market has a prescribed behavior that typically evolves in response to arbitrage pressures. Each futures market has a predetermined delivery charge. Due to arbitrage pressures, typically this basis converges to the delivery charge during the delivery option month.<sup>3</sup> Simply, due to arbitrage, it is expected that during the delivery period (or on the first day of delivery),  $B_d = DC_d$  where  $B$  is basis,  $DC$  is the delivery charge and  $d$  is a subscript referring to the delivery market. In periods prior to delivery, the basis may fluctuate, but generally conforms to the cost of storage from the period prior to delivery, until delivery occurs in which case the storage costs converge to zero.

In addition to this relationship, the intermonth price spreads (i.e., the difference in futures between successive delivery months) reflect the cost of storage and are impacted by the supply and demand for space. There is an elaborate theory explaining these market structures. Equilibrium intermonth price differentials are largely determined by the supply and demand for storage. This relationship is fundamental to most commodity market analysis.<sup>4</sup> Hieronymus summarizes these concepts:<sup>5</sup>

*...the nearby basis and spreads boil down to the supply and demand for space. When stocks at the terminal are large and grain is flowing to market rapidly, the cash price is weak in relation to the nearby futures and spreads are wide. But when stocks are small, the commodity is flowing to market slowly, and demand for shipment is vigorous, the price of storage decreases.*

And, Leuthold, Junkus and Cordier (1989) indicated:

*...the basis along with price spreads among futures contracts indicate the availability of commodity stocks. Large bases and price spreads represent either an abundance of stocks, or that the future market is providing an incentive to store the commodity for later release. Small or negative bases and price spreads associate with a shortage of stocks. The market is signaling for the release of stocks ....*

---

<sup>3</sup> The delivery process is described in Peck and Williams (1991), Coyle (2016), and Adjemian et al. (2013).

<sup>4</sup> This was illustrated in early writings on this topic including Working (1949) and followed by Gray and Peck (1981).

<sup>5</sup> Hieronymus (1971), op. cit. at 160.

There are a couple compounding factors that affect this general behavior. One is the introduction of the variable storage rate (VSR) for Chicago contracts. During the 2008-2010 time period, the cash price for SRW wheat in Toledo was significantly below the futures price at expiration (i.e., a weak basis) – more than \$2.00 at one point. This meant that commercial end-users of wheat for years could buy wheat for below the futures prices.<sup>6</sup>

The problem with the contract was ameliorated with two fixes to the CBOT SRW Wheat Futures Contract. One was an expansion of the number of delivery points. The other was implementation of the Variable Storage Rate (“VSR”) with the July 2010 contract, which corrected a design problem in the contract believed to be the cause of the convergence issues.<sup>7</sup> Briefly, the convergence problem was caused by a disconnect between the market cost of storage for wheat in the physical market and the storage rate for holding the delivery instrument, which has been set by the futures contract terms at an artificially low fixed rate.<sup>8</sup> This disconnect resulted in artificially narrow intermonth spreads and a weak basis at expiration. In response, and with industry consultation and CFTC approval, the CME created the VSR as a mechanism to encourage better convergence, implementing it with the July 2010 contract.

The VSR “is a market-based determinant of maximum allowable storage charges for outstanding wheat shipping certificate.” Greater maximum allowable storage rates are triggered to allow greater storage returns when intermonth spreads are wide. In addition, the maximum allowable storage charge is reduced when the inter-month spreads are narrow or inverted. After implementation of the VSR and other improvements to the contract, the performance of the contract improved significantly.

As is relevant here, there are several important points:

- In the period prior to adoption of the VSR mechanism, the intermonth spreads (which reflected returns to storage) were artificially constrained due to fixed storage rate in the contract.
- Implementing the VSR allowed intermonth spreads to expand to better reflect market rate of storage.
- This corrected the artificially low basis experienced in the period before the implementation of the contract fixes, resulting in a higher basis to commercial end-users such as Kraft.

Another exception is due to the prospect or potential for manipulation during the delivery process. The normal delivery process assumes a large number of buyers and sellers, that behave autonomously, and each are incapable of controlling delivery. Manipulation of this delivery

---

6 Seamon (2010) and Adjemian et al. (2013).

7 The VSR is described in detail at <http://www.cmegroup.com/trading/agricultural/grain-and-oilseed/variable-storage-rate.html>.

8 Adjemian et al. (2013), op. cit. at 8–16.

process can occur and may influence the basis at the delivery market. There are different forms of manipulation including what are referred to as a corner, squeeze or false report information (Kolb and Overdahl 2007). Further, manipulation may have an impact on the basis at the delivery market as described by Pirrong (2004). Finally, while manipulation is difficult to prove, there have been a number of examples including the Hunt silver case, Ferruzzi soybeans in 1989 (Hieronymus 1994b; Hieronymus 1994a), and more recently the alleged Kraft case on December 2011 CME wheat (Levine 2015). In these cases, the basis at the delivery market was impacted.

As it pertains to the scope of this study, the importance of the concepts above should be clarified. The concept of convergence is really focused only on the basis behavior at the delivery market. For other locations, there are a number of other factors that are important and do not necessarily adhere in any way to a normal converging pattern. For locations other than the delivery market, there are a number of important factors that vary temporally and spatially. These include spatial including international competition, discounts for quality, handling and shipping costs not all of which are fixed for locations away from the delivery location.

### ***Forecasting the basis:<sup>9</sup>***

There have been several recent studies that have addressed basis forecasting for storable commodities. Taylor et al. (2006) analyzed the wheat, corn and milo basis at six Kansas locations. The model was intended to be simple and the results indicated a historical one-year average model was optimal. In general, the naïve forecasts were more accurate for post-harvest basis projections. Hatchett et al. (2010) analyzed the basis in Oklahoma and Kansas for soft and hard wheat, corn and soybean. They sought to determine the moving average duration with the greatest accuracy in forecasting basis. If there is no structural change, the moving average was preferred, whereas if there is a structural change one should use the previous year's basis.

Time series modeling has been used in several studies to evaluate the forecasting performance of the basis. Onel and Karali (2014) used a semi-parametric, generalized additive model that accounts for nonlinearities in prices and basis to analyze forecasting. They applied it to weekly futures and soybean prices for corn and soybean in the North Carolina market. Three local origins were included for each, and they used data from 1988-2013. Their results indicated the semi-parametric models were preferred to traditional parametric time series models when forecasting the basis. Seamon, Kahl and Curtis, Jr. (2001) analyzed the cotton basis market for locational and seasonal differences in values. They pose that the basis should vary with transport costs, or at least regionally, as well as seasonally. They found that differences exist as expected, and the seasonal basis pattern is less pronounced in more westerly origins.

Sanders and Manfredo (2006) analyzed the soybean complex using time series methods (ARMA and VAR). Their data was from Central Illinois for each of soybean, soybean meal and oil from 1975 to 2004, with some of this used for out-of-sample evaluation. They concluded the simple

---

<sup>9</sup> The studies reviewed in this section are focused on grains and oilseeds. In addition to these, several studies have focused on non-storable commodities such as cattle (Tonsor, Dhuyvetter and Mintert 2004; Dhuyvetter et al. 2008).

5-year average basis might not be best. However, compared to other simple models, the improvement gained by time series is relatively small.

Three studies analyzed the spatial dimensions of basis behavior. Bekkerman, Brester and Taylor (2016) evaluate the hard wheat basis market across 215 origins in the upper Midwest to test the forecasting capabilities of numerous basis models. Important variables included specific market fundamentals, as well as the impact of the 2008 market year when there was an alleged market manipulation. In addition, elevator fixed effects were significant. They indicate this fixed effect is capturing ‘historical spatial and temporal basis relationships.’ Presumably, this is capturing in part the shipping cost and spatial competition in this market.

Manfredo and Sanders (2006) was of interest in that it sought to analyze the causality among export basis values and origin basis values. The model was applied to the US corn market and included export basis at Toledo and US Gulf, river terminals at Illinois River and Omaha and interior basis in Illinois, Iowa and Denver. The data were weekly and covered 1996-2005. Of interest, Toledo, as an export basis market, is dominant in discovery of basis values for other markets.

Finally, Welch, Mkrtchyan and Power (2009) analyzed the corn basis in the Texas Triangle. This study is of related importance for two reasons. One is that the Texas Triangle was used as a ‘deficit’ corn market. Second, the study includes an indicator of transport costs. Data were from 1997 to 2008. The transport cost indicator was an index and intended to capture impacts of oil price increases within the study period. Other variables included lagged basis, cash and futures prices ending stocks and several dummy variables. The transport cost variable was positive and significant. Results indicated this fundamental basis model performed better than a projection based on 3-year averages.

#### ***Factors affecting basis at non-delivery locations:***

A number of studies explored the behavior of the basis and to assess factors affecting its variability. An early study by Taylor and Tomek (1984) developed an econometric model to forecast the corn basis at a specific location in New York. They found that U.S. production, feed deficit in New York and open interest were each significant in explaining the basis. However, the model was limited in usefulness in that the explanatory variables were not easily predicable.

Parcell (2000) analyzed the impact of the Loan Deficiency Payment (LDP) program in Missouri on corn and soybean basis. Their econometric model includes lagged basis, futures, days to expiration, and location dummy variables, among others. Results indicated the LDP had only a minor impact on local basis values. Lara-Chavez and Alexander (2006) used an event study to assess impacts of Hurricane Katrina on the basis for corn, soybean and wheat. They found that Katrina had a larger impact on wheat futures than the other markets. Generally, they found an absence of abnormal returns in futures and basis for this market. They provided a number of explanations including that only logistics was affected, not the supply and demand for the commodities. Finally, Zhang and Houston (2005) analyzed how soybean production in South America and futures volatility impacted the basis. Though not exactly clear, it appears the basis



they analyzed is the “basis of the spot market located near the CBOT.” Their model found that both of these variables had a negative impact on the basis.

These studies relate to the current analysis in two respects. First, while Lara-Chavez and Alexander (2006) rejected the impacts of Katrina on markets, they did allude that the impacts may have impacted the transportation and logistics sector. Second, Zhang and Houston (2005) found that South American production had an important impact on the basis near the CBOT futures market. Our model focuses on each the impacts of transport and logistics on basis, as well as the impacts of Brazil on the export soybean basis.

### ***Export basis:***

Some of the studies above analyzed basis variability at specific locations away from the delivery market. This is of interest in that the current study analyzes basis values at export locations. Tilley and Campbell (1988) analyzed the US Gulf HRW basis. They found that the weekly variability was mostly explained by exports, free stocks, and the grain embargo. These were in addition to selected monthly variables included to capture seasonality in the basis. Notably shipping costs were not included in that analysis.

Another related and more recent study included all these variables, notable export and origin basis, in addition to shipping costs to evaluate spatial arbitrage opportunities in soybeans (Skadberg et al. 2015). Export locations at the US Gulf and PNW were included, and origin basis at a large number of interior origins. Shipping costs from each origin to destination included rail tariff rates, fuel service charges and second 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 copula distributions were important and accounted for the interrelations among basis values and shipping costs, implied in some other studies. Importantly, the model explicitly captures the relation between origin and destination basis, and shipping costs, in addition to spatial arbitrage and competition, as well as the interdependencies among these values, and risks. Several results are important. Origins in the upper Midwest have become highly dependent on the Pacific Northwest as a destination market. Second, arbitrage payoffs vary regionally. The results show how vertically integrated trading firms can capitalize on spatial-arbitrage payoffs. Finally, impacts of second car markets are illustrated.

### ***Impacts of rail (shipping costs) on basis, shippers and producers:***

Other researchers have studied the history and effects that the rail policy changes of the 1980s had on the grain markets. A number of studies have been conducted with the goal of examining the causal relationship between rail prices and basis levels/prices to producers. Wilson and Dahl (2011) was one of the first studies analyzing the interrelationships between basis and shipping costs. They found that basis values have become more volatile over time, and is impacted by factors such as shipping costs, ocean rate spreads, export sales, railroad performance, and others. In addition, they found that:

- 1) all marketing costs have increased;
- 2) increases in rail tariffs were less than those for barges;

- 3) secondary car market values, on average, declined;
- 4) fuel service charges had moderate changes in absolute terms; and
- 5) handling margins have had substantial increases, particularly at ports.

The econometric results indicate that the following variables were significant in explaining variability in origin basis values: shipping costs, Gulf-PNW ocean rate spreads, outstanding export sales, shipping industry concentration, rail performance (measured as cars late), and the ratio of stocks to storage capacity, futures prices, and varying measures of futures and destination spreads. These results validated other studies about increases in basis volatility and the importance of export sales' effect on basis. It also suggested performance in rail car shipments to be less of a determining factor in basis whereas other studies found the impact to be much greater.

Rail service disruptions caused by increased shipments from competing commodities, such as oil, have impacts on elevator prices. This was the focus of a number of studies. Olson (2014), in a study for Sen. Heidi Heitkamp, estimated that rail disruptions caused an aggregate loss to farmers statewide of \$66.6 million, or a little bit over \$2,000 per farm. This study did not analyze a direct relation between railroad price, performance, and basis. Rather it assumed that basis would be the same as an a priori determined analogue year, and then made derivations. In a report for the Minnesota Department of Agriculture, Usset (2014) used similar methods to estimate the impact of the 2013-2014 rail disruptions on Minnesota producers. Comparing 2014 to years with similar grain supply/demand levels, he estimated that farmers lost 40 cents/bushel on soybeans, 30 cents on corn, and 41 cents on hard red spring wheat. Another study by the USDA Office of the Chief Economist and Agricultural Marketing Service (2014) estimated the losses to be three percent of all farm cash receipts, but receipts but acknowledged the difficulty of pinpointing the exact cause of these losses. In another resulting study for the American Farm Bureau Federation, Kub (2015) further reviewed the 2013/2014 situation, but also argued that increasing infrastructure of truck, rail, barge, or pipeline transportation would reduce congestion of grain flow.

Villegas (2016) concluded that oil traffic, among other factors, is a determinant of wheat basis, and that this relationship is stronger in upper Midwest states, like North Dakota. The latest major example of these phenomena in the Upper Midwest was during the 2013-2014 marketing year when increased rail demand from oil and coal led to disruptions in grain shipping. Unable to move their inventory, shippers were forced to bid less aggressively for grain.

A more recent study was conducted by Hart and Olson (2017). Though it is really a study of origin basis behavior, it is included here as it seeks to address impacts of transport disruptions on local basis values. Corn, soybean and wheat basis patterns were analyzed in the major producing areas. Daily data were used for the period from 2003-2016. A model was estimated including export or terminal market basis in addition to indicators of ethanol and livestock production, the S&P, diesel, oceans shipping costs from the US gulf to Japan, the 2nd rail shuttle value, and indicators for months, winter, drought, and hurricane. Among other conclusions, the results showed that both ocean shipping costs and shuttle premiums were largely significant and had a negative impact on local basis values. The wheat market analysis was more complex and

included basis values at Portland and Minneapolis, export sales, rail tariff rates, 2nd rail shuttle values, among others. The results showed that export and terminal basis values, as well as shipping costs influence local basis values.

These studies motivated much discussion in the industry during these years. With the exception of Wilson and Dahl (2011) and Hart and Olson (2017), and for comparison to the current study, it is important that the other studies really conduct ‘analogue year’ analysis, using an a priori chosen year to be the analogue. Simply, the method is to choose a single similar year, and compare the market of the current 2013/14 year to the similar year. Second, none of these studies (except, as noted, Hart and Olson) included the cost of shipping in tariffs, fuel service charge (FSC) or secondary market car values, yet, they are seeking to determine the impact of shipping on basis values. While most previous studies focus on origin basis, they treat the export basis as exogenous and explanatory variable.

Third, there were several other very important facts concurrent with the study period included in these studies. Of importance, is that there was a large Brazil soybean crop in those years, and concurrently, a substantially improved logistical performance in that country. The impact of this was for downward pressure on port basis values in the US (PNW) that competes directly with Brazil in China and other markets. Each of these had impacts on port and local prices, notably at the country level. As example, in the case of corn, the PNW basis fell from +250c/b to less than 100c/b during this period. Solely looking at local basis would not capture this impact.

***Seasonal analog analysis (as applied to commodity prices and basis values):***

It is a commonly held belief that even though many commodity markets will exhibit a common seasonal pattern year-after-year, there can exist significant deviations from these seasonal patterns that are driven by market fundamental factors. For example, a recent article in *Real Money* (Garner 2015) observed:

*In most years, the market succumbs to these annual supply-and-demand factors to create the expected annual pattern. Nevertheless, these patterns are far from being a slam-dunk speculation. Along with the obvious market fundamentals, there are underlying factors influencing prices that we have yet to imagine, let alone understand.*

Kluis (2017), when comparing his seasonal charts from 10 years ago to current noted:

*When I updated my spreadsheets and created the new seasonal odds studies, I compared the current seasonal odds pattern with the seasonal odds pattern from 10 years ago. What I noticed was a big change in the corn seasonal odds pattern and a slight change in the soybean seasonal odds patterns.*

In commodity trading, these unique seasonal patterns are often grouped into what are called *seasonal analogs*. These analogs are typically grouped based upon a visual examination of seasonal year-on-year plots or through a rough application of correlation. Or the groupings are made based purely upon a particular fundamental factor or set of factors. For example, there

may be one analog based upon “large crop” years, and another for “small crop” years. There is the common adage that “short crops have long tails” and “long crops have short tails”.

The use of formal analog forecasting methods has a long history in meteorology and climatology research. Examples of recent research include Alexander et al. (2017) who used a new method referred to as “kernel analog forecasting” to forecast tropical oscillations. Djalalova, Delle Monache and Wilczak (2015) used ‘Kalman-filtering and analogs’ to evaluate among different forecasts of air quality. Finally, Comeau et al. (2017) used a prediction approach based on analog forecasting to analyze Arctic sea ice anomalies. The concept of analog years in weather were used to correlate planting date and corn yields (Elmore and Taylor 2013).

In agricultural economics, most of the research relating to analog seasonals has been focused upon relating weather analogs to crop yields and production (Hansen, Potgieter and Tippet 2004; Menzie 2007; Johansson et al. 2015; Irwin and Good 2016). Some extension publications, such as Flaskerud and Johnson (2000), published seasonal indices based upon crop fundamental analogs by grouping marketing years based upon a fundamental factor such as crop production. There have also been patents filed (Kolton, Gamboa and Chimenti 1996; Phillips et al. 2004) for systems using analog techniques to forecast commodity prices. Additionally, there have been recent studies examining the potential of analog techniques in the financial markets (Wanat, Śmiech and Papież 2016; Lahmiri, Uddin and Bekiros 2017).

Bullock (2004) applied a multivariate technique called *exploratory factor analysis* (EFA) to derive seasonal analogs for Minnesota hard red spring wheat prices and basis values (September and July futures) for the 1960/61 through 2003/04 marketing years. For price, a total of 8 unique analogs were derived with two unique outlier years (1973/74 and 1974/75). A total of 6 and 5 analogs were identified for the September and July basis values. Using fundamental balance table data, Bullock tried to correlate the seasonal price analogs using multinomial logit, stepwise OLS, and single equation probit models but found the results to be very poor in terms of statistical fit (low  $R^2$ ) and out-of-sample forecasting ability.

### ***Summary Observations:***

A few observations from these studies are important as it relates to the current study. For locations other than the delivery market, there are a number of important factors that vary temporally and spatially. These include spatial competition, in addition to international competition, discounts for quality, handling and shipping costs not all of which are fixed for locations away from the delivery location. It is important these be captured in any evaluation of non-delivery basis values. There are several important points.

One is that most of these studies are largely domestic and focused on origin basis. In some cases, these use the export basis as an explanatory variable for the origin basis. Only a couple of the studies focuses on a destination basis, or export basis.

Second, for basis values away from the delivery market, a number of other factors are critically important. These include notably, impacts of shipping costs both intra and inter-country. With exception of Hart and Olson (2017), these are largely ignored or proxied using indices more

generic. Third, some of the studies sought to evaluate the impacts of shipping disruptions on basis values. In some cases, these were done by analog year comparisons. In those cases, shipping costs were not measured, but instead were captured via the assumed analog year comparison. In Hart and Olson (2017), they were captured from the second market value.

## **Data and Methodology**

### ***Conceptual Model***

The basis is defined as the difference between a particular spot or forward contract price and the futures price. The futures price represents a highly standardized market; therefore, the basis represents the adjustments necessary to convert the standardized price into the non-standardized spot or forward contract market price. For a storable commodity, such as soybeans, the basis is composed of four basic components: (1) the cost of carry from the cash market maturity to the futures contract maturity, (2) the cost of transporting the commodity from the cash market to the futures par delivery market, (3) a quality premium or discount accounting for the difference between the cash market commodity and the defined par delivery commodity, and (4) a random component that includes sufficient adjustments to assure a sufficient flow of the commodity to meet the demand at the relevant cash market. The supply / demand influence of the fourth component is sometimes called the *pipeline* component and it reflects the coordination role that is vital to a well-functioning commodity market.

The basis at export markets responds to meet competition (competitive parity) from two sources: (1) *international* competition from competing exporters of the commodity, and (2) *internal* competition from domestic users of the commodity. For U.S. soybeans, the primary international competition comes from the South American producing countries (primarily Brazil). For internal domestic competition, the soybean export basis must be great enough to draw origin soybean production away from the domestic crushing industry to the export channels. Soybeans are not stored at export locations; therefore, futures carry is more an indicator of current versus future demand. A strong inverse carry is an indicator of strong nearby demand; therefore, the export basis must strengthen in order to assure a sufficient flow of soybeans into the export channels via domestic market channels.

Export basis must be great enough to attract soybeans to the export location and away from the par futures delivery market and other domestic markets. These costs include rail shipping (tariff plus fuel surcharge), and barge rates. The export basis may also be influenced by ocean shipping costs – for soybeans, primarily to the Asian markets. Higher ocean freight costs are hypothesized to put downward pressure on export basis values to maintain competitive parity in the international market. However, this essentially depends upon the relative change in ocean shipping costs between the U.S. export locations and international competitors.

Previous research (Wilson and Dahl 2011; Hart and Olson 2017) indicated that logistical conditions, such as railcar shortages and secondary railcar market values, can have significant impacts upon origin basis values. These variables may also have an impact upon the export basis and are therefore included in this analysis. Additionally, measures of current and anticipated

export activity such as export inspections, ships in port, and outstanding export commitments may have a direct impact upon export basis values and are included in the analysis.

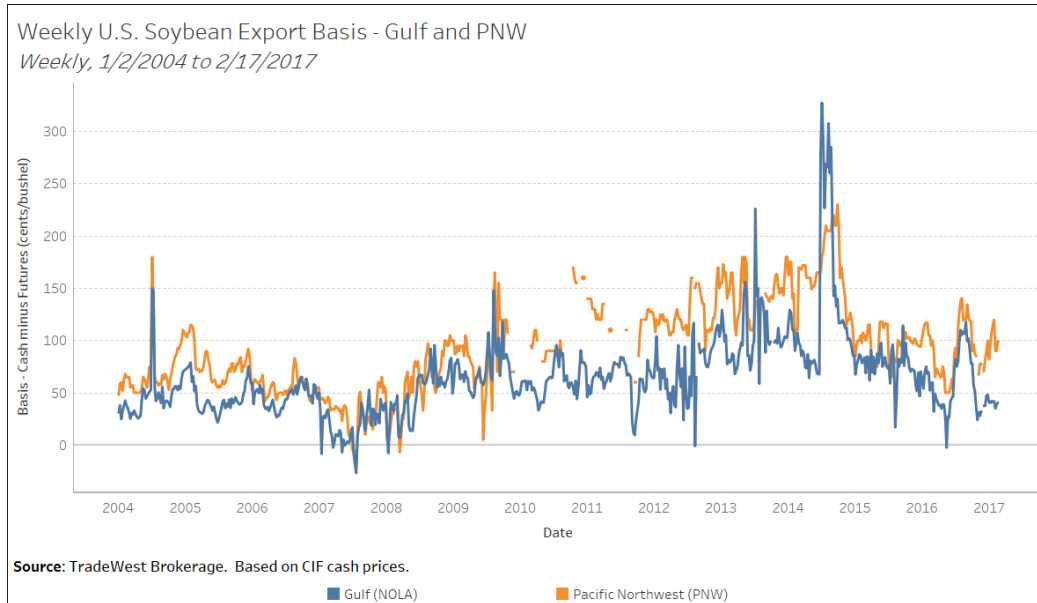
The following conceptual models are proposed and examined in this study:

$$\begin{aligned}\bar{B}_{i,t} &= f(IntD_t, DomD_t, Trans_t, Logistic_t, ExpActivity_t), \text{ and} \\ S(B_{i,t}) &= f(IntD_t, DomD_t, Trans_t, Logistic_t, ExpActivity_t),\end{aligned}\tag{1}$$

where  $i$  is a subscript for the particular export market (1 for Gulf, 2 for PNW),  $t$  represents the marketing year (2004/05 through 2015/16),  $\bar{B}$  represents the marketing year average export basis value,  $IntD$  is a set of variables representative of international demand and competition,  $DomD$  is a set of variables representative of domestic demand and competition,  $Trans$  is a set of variables representing transportation costs,  $Logistic$  is a set of variables representing logistical conditions,  $ExpActivity$  represents a set of variables representing levels of export activity, and  $S(\cdot)$  represents a transformation of the monthly basis values into a particular seasonal analog.

### ***Dependent Variables***

Weekly nearby basis data (CIF delivered rail) from 1/2/2004 through 2/17/2017 for both the Gulf (NOLA) and Pacific Northwest (PNW) export markets was obtained from TradeWest Brokerage and is illustrated in Figure 1. Missing values (64 total) were interpolated using the NIPALS (Wold 1973) procedure. The data was then converted into monthly and marketing year (September through August) averages to be used in the analyses to follow.



*Figure 1*

## Independent (Explanatory) Variables

A complete list of the potential explanatory variables, sources, and levels of aggregation is given in Table 1. All of the listed variables are rolled up to a marketing year (MY) average or value for the analyses to follow. Freemont, NE was chosen as the interior point for the rail costs due to its position relative to the Gulf and PNW markets as a tributary shipper to both markets.

*Table 1. List of Explanatory Variables Used in the Analyses*

Category	Variable Name	Description	Source	Aggregation
<b>International Demand (IntD)</b>	<i>Basis-Brz</i>	Brazil Soybean Basis, Paranagua - CME Futures (US\$/bu)	CME, Cepea	MY Average of Weekly Values
	<i>SA-Prod</i>	Total South American (Argentina-Brazil-Paraguay) Soybean Production (mmt)	USDA - WASDE	MY Reported Values
	<i>China-Import</i>	China Total Soybean Imports (mmt)	USDA - WASDE	MY Reported Values
	<i>World-SU</i>	World Soybean Ending Stocks-Use Ratio (%)	USDA - WASDE	MY Reported Values
<b>Domestic Demand (DomD)</b>	<i>Futures-NB</i>	Nearby CBOT Soybean Futures Price (c/bu)	CME, DTN ProphetX	MY Average of Weekly Values
	<i>FutSprd1</i>	2nd NB Futures - NB Soybean Futures Price (c/bu)	CME, DTN ProphetX	MY Average of Weekly Values
	<i>FutSprd2</i>	3rd NB Futures - 2nd NB Soybean Futures Price (c/bu)	CME, DTN ProphetX	MY Average of Weekly Values
	<i>SU-Ratio</i>	U.S. Soybeans Ending Stocks-Use Ratio (%)	USDA - WASDE	MY Reported Values
	<i>MealP</i>	U.S. Domestic Soybean Meal Price (\$/ton)	USDA - WASDE	MY Reported Values
	<i>OilP</i>	U.S. Domestic Soybean Oil Price (c/lb)	USDA - WASDE	MY Reported Values
<b>Transportation Costs (Trans)</b>	<i>Crush</i>	U.S. Domestic Soybean Crush as % of Total Supply	USDA - WASDE	MY Reported Values
	<i>Rail-Gulf</i>	Total Rail Cost, Shuttle Trains (Tariff plus Fuel Surcharge) - Freemont, NE to Texas Gulf (\$/car)	BNSF Railroad	MY Average of Weekly Values
	<i>Rail-PNW</i>	PNW versus Gulf Rail Cost Spread, Shuttle Trains - Freemont, NE (\$/car)	BNSF Railroad	MY Average of Weekly Values
	<i>Barge-Spot</i>	Spot Barge Rate - St. Louis to Gulf (\$/ton)	USDA - AMS	MY Average of Weekly Values
	<i>Barge-3M</i>	3-month Forward Barge Rate - St. Louis to Gulf (\$/ton)	USDA - AMS	MY Average of Weekly Values
	<i>Ocean-PNW</i>	Ocean Freight Rate from PNW to Japan (\$/mt)	USDA-AMS	MY Average of Weekly Values
<b>Logistic Conditions (Logistic)</b>	<i>Ocean-Sprd</i>	Ocean Freight Spread to Japan - Gulf versus PNW (\$/mt)	USDA-AMS	MY Average of Weekly Values
	<i>Cars-Late</i>	Average BN Railcars Placed Late (# of cars)	BNSF Railroad	MY Average of Weekly Values
	<i>DCV</i>	Daily Secondary Market Car Values for Shuttle Trains (\$/car)	TradeWest Brokerage	MY Average of Weekly Values
	<i>FarmDel-Q1</i>	Cumulative farmer delivery % through Q1 of MY	USDA - NASS	Cumulative Sum of Monthly
	<i>FarmDel-Q2</i>	Cumulative farmer delivery % through Q2 of MY	USDA - NASS	Cumulative Sum of Monthly
	<i>FarmDel-Q3</i>	Cumulative farmer delivery % through Q3 of MY	USDA - NASS	Cumulative Sum of Monthly
<b>Level of Export Activity (ExpActivity)</b>	<i>Gulf-InPort</i>	Average Number of Ships in Gulf Ports (ships)	USDA - AMS	MY Average of Weekly Values
	<i>PNW-InPort</i>	Average Number of Ships in PNW Ports (ships)	USDA - AMS	MY Average of Weekly Values
	<i>Export-Gulf</i>	FGIS Export Inspections at Gulf Ports (1000 bushels)	USDA - FGIS	MY Average of Weekly Values
	<i>Export-PNW</i>	FGIS Export Inspections at PNW Ports (1000 bushels)	USDA - FGIS	MY Average of Weekly Values
	<i>Export-Out</i>	U.S. Soybean Export Sales - Outstanding Balance (1000 bushels)	USDA - FAS	MY Average of Weekly Values

## Methodology

The dataset contains 2 dependent variables and 27 potential explanatory variables observed over 12 marketing years. Due to the overidentification of the explanatory variable matrix, ordinary least squares (OLS) regression cannot be applied. Additionally, there exists a high degree of correlation between individual variables in the explanatory dataset which indicates issues with multicollinearity.

To address both of these issues, the partial least squares (PLS) regression model (Wold 1966) was used to estimate the impact of the explanatory variable set upon the market year average basis for both the Gulf and PNW. PLS regression is particularly useful when predicting a set of dependent variables from a large set of independent variables (Abdi 2007) and has found numerous applications in chemometrics (Wold 2001) and sensory evaluation (Martens and Naes 1989). It has also achieved greater recent popularity in the social sciences as a multivariate tool for examining both non-experimental and experimental data within a structural equation (PLS-SEM) framework (F. Hair Jr et al. 2014).

Traditionally, finding solutions to regression problems in the presence of multicollinearity and/or data sparsity follows one of two common approaches: (1) removal of the highly correlated predictors through one of a variety of techniques, or (2) conducting principal component analysis (PCA) on the explanatory variables and regressing the dependent variables upon the extracted

principal components (Kuhn and Johnson 2013, p. 113). The latter technique is known as principal component regression (PCR, Massy 1965) and has been one of the most commonly used procedures in social science research. One primary disadvantage of PCR is that it is an *unsupervised* procedure – it only considers information in the explanatory variables when constructing the components. If variability in the explanatory variable space is not related to the variability in the dependent variable, then PCR has a high probability of not correctly identifying all the predictive relationships in the data.

PLS, on the other hand, extracts its components (called *latent variables*) in the direction of optimizing the covariance between the dependent variable(s) and the explanatory variables. Therefore, it is a *supervised* procedure in that it considers information in both the dependent and explanatory variable sets. PLS also has the advantage of considering multiple dependent variables since the components are directed towards explaining the covariance of the dependent (Y) and explanatory (X) data matrices.

The general underlying model for multivariate PLS is as follows (Esbensen and Swarbrick 2018) where the explanatory (X) and dependent (Y) variables are simultaneously decomposed:

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \quad (2)$$

where  $X$  is an  $n \times m$  matrix of explanatory variables,  $Y$  is an  $n \times p$  matrix of dependent variables,  $T$  and  $U$  are  $n \times r$  matrices that contain the  $X$  and  $Y$  scores respectively,  $P$  and  $Q$  are  $m \times r$  and  $p \times r$  orthogonal *loading* matrices, and  $E$  and  $F$  represent the error terms which are independent and identically distributed random normal variables. These decompositions of  $X$  and  $Y$  are conducted with the goal of maximizing the *covariance* between the  $T$  and  $U$  matrices. The regression coefficients for the latent variables can then be derived by finding an  $r \times r$  matrix of regression coefficients ( $\beta_{PLS}$ ) such that:

$$U = T\beta_{PLS} \quad (3)$$

Noting that equation (1) implies that  $T = XP$  and therefore, the individual PLS regression coefficients for each explanatory variable are the sum product of the variable's loadings (from  $P$ ) and the PLS latent variable regression coefficients ( $\beta_{PLS}$ ).

Given that PLS results are highly sensitive to the unit measurements of the explanatory variables, it is generally a practice to convert all variables to a *z-score* equivalent by normalization prior to estimation of the model. The number of retained components (latent variables) in PLS is generally guided by a cross-validation procedure using a goodness-of-fit or goodness-of-prediction metric such as the proportion of variance explained in the  $Y$  matrix ( $R^2Y$ ) or the out-of-sample predictability of the model using predicted-residual-sum-of-squares (PRESS) or the  $Q^2$  statistic.

A number of procedures are available to determine the optimal scoring and loading matrices for PLS regression, the most popular being the NIPALS algorithm originally developed by Wold



(1973). The implementation used in this paper is the PLS-R procedure from the *XLStat*<sup>10</sup> software package which is based upon the NIPALS algorithm.

One very useful output of the PLS regression model is an index assigned to each explanatory variable called the *variable importance in projection* (VIP) which can be expressed by the following equation (Wold et al. 1993):

$$VIP_j = \sqrt{\frac{\sum_{i=1}^h R^2(y, t_i) (w_{ij} / \|w_i\|)^2}{(1/p) \sum_{i=1}^h R^2(y, t_i)}} \quad (4)$$

where  $p$  is the number of predictor variables,  $h$  is the number of retained component variables,  $w_{ij}$  is the weight of the  $j$ -th predictor variable in component variable  $i$ , and  $R^2(y, t_i)$  is the fraction of variance in  $Y$  explained by component  $i$ . The VIP score represents the proportion of explained variance of  $X_j$  relative to  $Y$  (through the component variables) divided by the average explained variance between all  $X$  variables and  $Y$ . By definition, the average squared VIP score is equal to one; therefore, a typical rule for variable selection is to retain all variables whose VIP score is greater than one.

To analyze basis seasonality, the monthly basis values were converted into additive seasonal indices by subtracting the monthly basis value from the marketing year average. Marketing year basis patterns were then grouped into similar seasonal pattern analogs by applying the agglomerative hierarchal clustering (AHC, Ward 1963) to the index data.

The AHC procedure is an iterative classification method that starts by calculating the dissimilarity in the seasonal patterns between the 12 marketing years. The proximity between marketing years is measured using the Euclidian distance metric. The first two marketing years are clustered together based upon the minimization of Ward's agglomeration criterion which aggregates so that the within-group inertia increases as little as possible. Then the iterative process continues by calculating the dissimilarity between this first class the remaining 10 marketing years. This process continues until all the objects have been clustered together.

The successive clustering operations produce a binary clustering tree called a *dendrogram* whose root is the class that contains all the marketing years. This graphically represents the hierarchy of partitions. The final set of analogs are chosen by truncating the dendrogram at a determined level using a specified criterion. In this study, the minimum entropy criterion is used to determine the point of truncation. The measurement of entropy uses an index originally developed by Shannon (1948).

To characterize the individual seasonal analogs, an two-sample z-test originally proposed by Lebart et al. (2000) was applied. The test statistic is similar to a classic t-test but is applied to

---

<sup>10</sup> Addinsoft (2019). XLSTAT statistical and data analysis solution. Boston, USA. <https://www.xlstat.com>

test the difference in means between a particular set and subset of observations. Since the two estimates are correlated, a classic t-test cannot be used. The test statistic is:

$$z_k(X) = \frac{\bar{X}_k - \bar{X}}{s_k(X)},$$

$$\text{where } s_k^2(X) = \frac{n - n_k}{n - 1} \cdot \frac{s^2(X)}{n_k},$$
(5)

with  $\bar{X}_k$  and  $\bar{X}$  equal to the sample means from the subset and parent set,  $n_k$  and  $n$  equal to the number of observations in the subset and parent set, and  $s^2(X)$  equal to the parent set variance. The  $z_k$  statistic is distributed as asymptotically unit normal so it can be treated as a standard z-statistic for determining statistical significance.

## Results

### *Statistical Characterization of the Market Year Average Basis for Gulf and PNW*

The marketing year average basis levels for Gulf and PNW soybean export markets had a noticeable shift higher in the mean MY basis level beginning with the 2008/09 marketing year for both markets. Therefore, for the analysis of the marketing year averages, the explanatory dataset (Table 1) is augmented with a dummy variable (*Prior to 2008?*) which is equal to one if prior to 2008/09 and a linear trend (*Trend*) variable. The reasons for these are due in part to the radical change and increase in volatility in all commodity markets following the 2008/09 crop year, and the growing trend in U.S. soybean exports, particularly to China.

The mean values from 2004/05 through 2007/08 were 38 and 55 cents per bushel for the Gulf and PNW respectively. For 2008/09 through 2015/16, the average basis levels were 80 and 114 cents per bushel respectively. A two-sample one-tail t-test with the null hypothesis of no difference between the means versus the alternate hypothesis that the mean in the first period was less than the latter was applied. The t-test rejected the null hypothesis in favor of the alternative hypothesis at the 99% confidence level for both the Gulf and PNW markets.

The volatility of the basis has also escalated over time. The standard deviation of the basis (derived using monthly data) for the period prior to 2008/09 was 9.53 and 18.58 cents per bushel respectively for the Gulf and PNW. For the period following, these values increased to 24.03 and 26.72 cents per bushel for the Gulf and PNW respectively. Application of Fisher's F-test to the basis variance ( $H_a$ : first period variance is less than latter period), however, indicated that only the Gulf basis was significantly lower (at 90% confidence level) while the PNW was not significant at the 90 percent level or higher.

The MY average basis values for the Gulf and PNW are highly correlated (93.0%) over the 12 market year observations; therefore, they were both included in the  $Y$  matrix for the PLS-R estimation. For the analysis to follow, a two-part procedure was used. First, the PLS-R model was applied to the full explanatory dataset. Those variables whose variable importance in

projection (VIP) indices were greater than one will be retained in the dataset and those with values less than one are removed. The second part applied PLS-R to the reduced dataset for the final estimation of the regression coefficients and interpretation of significance. The number of PLS-R components to retain was determined by cross-validation using the *jackknife leave-one-out* (Jackknife-LOO) procedure. This procedure sequentially leaves each observation out of the estimation set for validation testing. The number of components that maximizes the  $Q^2$  quality index statistic was chosen.

Application of the PLS-R procedure to the full explanatory dataset resulted in one retained component (latent variable) with an optimal  $Q^2$  statistic of 0.6555 using the Jackknife LOO procedure. The value of the  $Q^2$  statistic indicates that the retained latent variable explains approximately 65.5% of the variability in the out-of-sample values for the Y matrix containing the MY average basis levels for the Gulf and PNW markets. The  $R^2Y$  was equal to 0.788 which indicates that the one retained latent variable explains approximately 78.8% of the in-sample variability in the Y matrix.

Less than half (13 out of 29) of the variables had a VIP score greater than one and were retained for the final regression estimation procedure (Figure 2) and are shaded in red. The results indicate that the Brazilian export basis (*Basis-Brz*) is the most important variable in projecting both the Gulf and PNW basis levels. Following in importance are the nearby futures spreads (*FutSprd1* and *FutSprd2*), the domestic soybean meal price (*MealP*), and the total rail shipping costs from Fremont, NE to the Gulf (*Rail-Gulf*). The additional variables retained (in order of their VIP scores) were *Prior to 2008?*, *PNW-InPort*, *Export-Out*, *Trend*, *China-Import*, *SU-Ratio*, *Rail-Sprd*, *Futures-NB*, *SA-Prod*, *Export-Gulf*, *DCV*, *FarmDel-LQ2*, *Ocean-PNW*, *Gulf-InPort*, *Export-PNW*, *Crush*, *FarmDel-LQ1*, *World-LSU*, *FarmDel-LQ3*, *Cars-Late*, *OilP*, *Barge-Spot*, *Barge-3M*, and *Ocean-Sprd*.

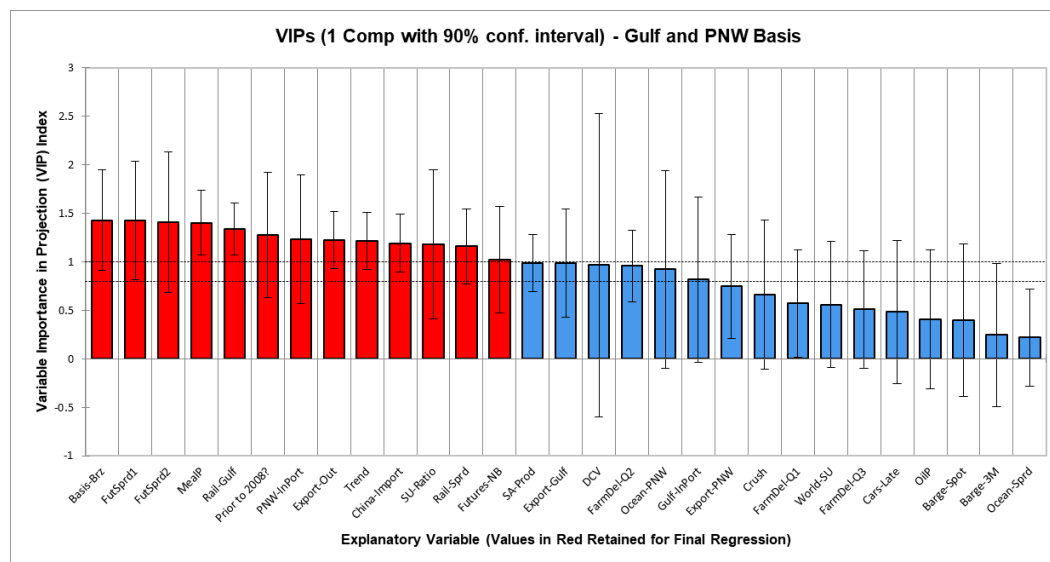


Figure 2

The second round of PLS-R estimation regressed the Y matrix of MY average basis values for both the Gulf and PNW upon the one retained latent variable component derived from optimizing the covariance between the Y matrix and the VIP reduced explanatory variable set (X

matrix). This resulted in a quality index ( $Q^2$ ) statistic of 0.7198 which indicates a substantial improvement in the out-of-sample predictability of the model (6.43% gain in  $Q^2$  from the full to the reduced variable set). The in-sample  $R^2Y$  equaled 0.798 which indicates the latent variable accounts for almost 80% of the variability in the Y matrix containing the Gulf and PNW average basis levels.

Table 2 shows the PLS-R regression equation for the Gulf marketing year average basis level. The individual equation regression fit had an  $R^2$  coefficient of 0.7989 with a root mean squared error (RMSE) of 12.3 cents per bushel (both in-sample). The non-standardized and standardized coefficient estimates are both presented with the coefficients ordered by the absolute value of the standardized coefficient estimates.

*Table 2. PLS-R Regression Results for Gulf Market Year Average Basis Level*

Variable	Coefficient	Standardized Coefficient	Std. deviation	Lower bound (90%)	Upper bound (90%)	T-Statistic	Significance
Intercept	-1.4753	N/A	14.7704	-28.0012	25.0506	-0.0999	0.9222
Basis-Brz	0.0722	0.0920	0.0243	0.0286	0.1159	2.9729	0.0127 **
FutSprd1	-0.1804	-0.0917	0.0695	-0.3053	-0.0555	-2.5944	0.0249 **
FutSprd2	-0.1668	-0.0905	0.0715	-0.2952	-0.0383	-2.3310	0.0398 **
MealP	0.0256	0.0902	0.0063	0.0142	0.0370	4.0370	0.0020 ***
Rail-Gulf	0.0042	0.0862	0.0008	0.0029	0.0056	5.6118	0.0002 ***
Prior to 2008?	-4.7869	-0.0822	0.9871	-6.5596	-3.0143	-4.8497	0.0005 ***
PNW-InPort	0.8422	0.0792	0.3737	0.1710	1.5134	2.2535	0.0456 **
Export-Out	0.0240	0.0789	0.0054	0.0143	0.0337	4.4286	0.0010 ***
Trend	0.6231	0.0783	0.0994	0.4446	0.8015	6.2695	0.0001 ***
China-Import	0.1122	0.0766	0.0207	0.0749	0.1494	5.4080	0.0002 ***
SU-Ratio	-0.4607	-0.0758	0.1698	-0.7656	-0.1559	-2.7140	0.0201 **
Rail-Sprd	0.0094	0.0745	0.0026	0.0048	0.0141	3.6269	0.0040 ***
Futures-NB	0.0063	0.0656	0.0017	0.0032	0.0094	3.6482	0.0038 ***

The results indicate that all of the explanatory variable coefficients are statistically significant at the 95% level or greater. The Brazilian MY average export basis value in US\$ per bushel (*Basis-Brz*) has the greatest individual impact upon the Gulf average basis value and the coefficient estimate is statistically significant at the 95% confidence level. The sign of the coefficient is positive which indicates that maintaining international competitive parity with Brazil is the most important factor in setting the Gulf export basis value.

The nearby futures carry spreads (*FutSprd1* and *FutSprd2*) are next in terms of statistical importance with negative signs for both coefficients. Strong negative carry in the intermonth futures spreads is an indication of strong current demand for soybeans when compared to demand in the deferred months. In order to draw stocks of soybeans into the export channel to meet the higher current demand, a higher basis level is necessary in order to maintain market competitiveness for the current supply of soybeans.

The domestic soybean meal price (*MealP*) is fourth in importance, has a positive coefficient value, and is significant at the 99% confidence level. The sign is as expected since higher soymeal prices would indicate increased domestic crush demand for soybeans. The export basis levels would have to adjust higher in order to maintain competitive parity with the domestic crush demand.

The top four explanatory variables (*Basis-Brz*, *FutSprd1*, *FutSprd2*, and *MealP*) are all reflective of competitive demand pressures that are applied to the Gulf export market; therefore, this

supports the hypothesis that a role of the Gulf export basis level is to respond to competitive pressures, both from abroad (*Basis-Brz*) and domestically (*FutSprd1*, *FutSprd2*, and *MealP*).

The only logistical cost factors in the reduced explanatory variable set are the costs of rail transportation (sum of tariff and fuel surcharge) from Freemont, NE to the Gulf (*Rail-Gulf*) and the additional cost to the PNW (*Rail-Sprd*). Both of these costs are significant at the 99% level and positive in sign. The cost to the Gulf (*Rail-Gulf*) has a higher standardized coefficient value and t-statistic when compared to the additional cost to the PNW (*Rail-Sprd*). The positive sign on the rail cost to the Gulf is expected since the Gulf export basis must adjust higher in order to maintain competitiveness with the domestic market. However, the positive sign on the additional cost to the PNW is not expected as the logic would follow that Gulf basis levels would not have to increase as much to maintain competitive parity with the PNW. The only potential explanation would be that the more important rail cost is the total cost to the PNW (equals *Rail-Gulf* + *Rail-Sprd*) rather than to the Gulf, where barge shipments are also a major part of the logistics. There is also evidence that export activity out of the PNW is more important to overall basis levels in both locations as the average number of ships in port at the PNW (*PNW-InPort*) is in the reduced explanatory variable set while the average number of ships in port at the Gulf (*Gulf-InPort*) did not make the initial cut.

Both the early period dummy variable (*Prior to 2008?*) and the market year trend (*Trend*) variables were highly significant (99% confidence level) and had the anticipated signs (negative for period dummy and positive for trend). These are important results confirming that the basis is greater in the period following 2008.

The average number of ships in port at the PNW (*PNW-InPort*) was significant at the 95% level and had the expected positive sign reflecting a larger volume of export activity at the PNW. The absence of the similar Gulf measure indicates that activity in the PNW is more important as an influence upon overall basis values at both locations and reflective of the emerging primacy of the PNW market with the emergence of China as a major exporter of soybeans.

The average level of outstanding export sales (*Export-Out*) is highly significant (99% level) and the positive sign of the coefficient is as expected since higher anticipated export demand should be reflected in higher export basis values in order to move stocks into position to meet anticipated demand. The volume of total soybean imports by China (*China-Import*) is highly significant (99% confidence level) and the positive sign is as expected since China is the number one source of export demand for U.S. soybeans over the past decade. Even though a significant portion of this demand moves through the PNW export market, the increase in overall export demand also has a positive effect upon the Gulf demand and basis level also.

The domestic soybean ending stocks-use ratio (*SU-Ratio*) made the initial cut while the global stocks-use ratio (*World-SU*) did not. The coefficient estimate is significant at the 95% level and the negative sign of the coefficient is as expected. A higher domestic stocks-use ratio indicates that demand is low relative to existing supply; therefore, the export basis level can decline and still maintain competitive parity with the domestic market demand. Also, the significant (99%) and positive sign with the MY average nearby futures price level indicates that basis levels will

generally adjust in the same direction as the overall level of demand for soybeans relative to supply.

Table 3 shows the PLS-R regression results with the PNW marketing year average basis as the dependent variable. The individual equation regression fit had an  $R^2$  coefficient of 0.7971 with an RMSE of 16.0 cents per bushel.

*Table 3. PLS-R Regression Results for PNW Market Year Average Basis Level*

Variable	Coefficient	Standardized Coefficient	Std. deviation	Lower bound (90%)	Upper bound (90%)	T-Statistic	Significance
Intercept	6.7783	N/A	14.8033	-19.8067	33.3632	0.4579	0.6559
Basis-Brz	0.0934	0.0919	0.0306	0.0385	0.1482	3.0558	0.0109 **
FutSprd1	-0.2331	-0.0916	0.0838	-0.3836	-0.0827	-2.7830	0.0178 **
FutSprd2	-0.2155	-0.0904	0.0854	-0.3689	-0.0621	-2.5229	0.0283 **
MealP	0.0331	0.0901	0.0068	0.0209	0.0453	4.8700	0.0005 ***
Rail-Gulf	0.0055	0.0861	0.0008	0.0040	0.0069	6.8680	0.0000 ***
Prior to 2008?	-6.1860	-0.0821	1.5748	-9.0140	-3.3579	-3.9282	0.0024 ***
PNW-InPort	1.0883	0.0791	0.4536	0.2738	1.9029	2.3995	0.0353 **
Export-Out	0.0310	0.0788	0.0062	0.0198	0.0422	4.9644	0.0004 ***
Trend	0.8052	0.0782	0.1205	0.5888	1.0215	6.6824	0.0000 ***
China-Import	0.1450	0.0766	0.0264	0.0976	0.1923	5.4979	0.0002 ***
SU-Ratio	-0.5954	-0.0757	0.2305	-1.0094	-0.1814	-2.5829	0.0255 **
Rail-Sprd	0.0122	0.0745	0.0025	0.0076	0.0167	4.8195	0.0005 ***
Futures-NB	0.0081	0.0655	0.0020	0.0044	0.0118	3.9777	0.0022 ***

Given the very high level of Pearson correlation (93%) between the MY average basis values for both the Gulf and PNW, the regression results for the PNW are nearly similar to the Gulf with a few exceptions. First, the intercept for the PNW equation is 6.7783 while for the Gulf it is -1.4753 indicating an average implied fixed premium of a little over 8 ¼ cents per bushel over the 12-year study period. However, the high standard deviations for both coefficients indicate that the difference is not statistically significant.

Second, the coefficient estimates for the PNW equation have a higher magnitude in value but match the signs of the Gulf equation, indicating a slightly higher impact for each variable. However, this difference in impact can almost be completely attributed to a higher variability in the PNW basis as the standardized coefficients are nearly identical between the two estimates. When examining the loadings of the two export basis vectors upon the one retained latent variable, the Gulf (0.7989) had a higher loading when compared to the PNW (0.7971). However, from an out-of-sample forecasting perspective, using the Jackknife-LOO cross-validation procedure, the PNW had a slightly higher quality ( $Q^2$ ) index value of 0.7263 when compared to the Gulf (0.7134).

The coefficient standard errors and t-statistics for the PLS-R estimation procedure are estimated directly from the cross-validation procedure; therefore, the coefficient estimated for the PNW have slightly higher (in absolute value) t-statistics when compared to the Gulf.

From these results, the following observations can be made. First and foremost, factors influencing the overall marketing year average level of the export basis are similar between the Gulf and PNW in terms of importance ranking, statistical significance, and level of impact. The primary factors influencing the average basis level are competition from Brazil (*Basis-Brz*) and the domestic market (*FutSprd1*, *FutSprd2*, and *MealP*). Export basis adjusts to competitor basis

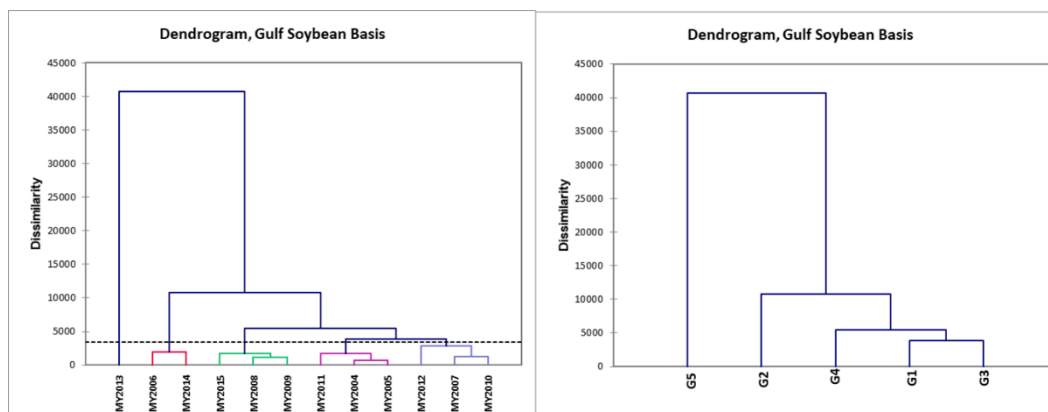
as above which affects international competitiveness. It also influences the amount of stocks in exportable position to meet international demand, particularly from China.

Second, the rise of the Chinese export demand has increased the relative importance of the PNW market over time despite the fact that the percentage of exports moving through the PNW is still less than the volume moving through the Gulf. Activity directly related to the volume of exports out of the PNW (*PNW-InPort*) has more influence upon the MY average basis levels at both the Gulf and PNW when compared to a similar measure of activity (*Gulf-InPort*) out of the Gulf.

Third, internal logistical costs are of secondary importance and are primarily limited to rail costs (tariff and fuel surcharge). The export basis level must respond positively with these costs in order to maintain parity with domestic demand and assure adequate export market flows and supplies.

### ***Seasonal Analog Derivation***

Application of AHC clustering to the Gulf soybean data produces the dendrograms shown in Figure 3. The left-hand side shows the application of the Euclidian distance and Ward's agglomeration to the marketing years. The dashed line shows the minimum entropy cut-off point for selecting the analogs. This results in five distinct seasonal analog groupings that are shown in Table 4.



*Figure 3. Dendrograms for application of AHC to Gulf soybean basis seasonal indices.*

Note that no single analog appears to be dominant with analogs G1, G2, and G4 containing three marketing years; analog G2 containing two marketing years, and analog G5 containing a single marketing year.

Table 4. Summary of AHC Seasonal Analog Groupings for Gulf Basis

Analog	G1	G2	G3	G4	G5
Objects	3	2	3	3	1
Sum of weights	3	2	3	3	1
Within-class variance	1223.42	1954.39	2044.86	1409.71	0.00
Minimum distance to centroid	25.12	31.26	30.54	26.49	0.00
Average distance to centroid	28.23	31.26	36.54	30.52	0.00
Maximum distance to centroid	34.36	31.26	43.39	33.36	0.00
	MY2004	MY2006	MY2007	MY2008	MY2013
	MY2005	MY2014	MY2010	MY2009	
	MY2011		MY2012	MY2015	

Figure 4 shows a plot of the average seasonal indices for each of the five Gulf seasonal analogs. Analog G1 shows what would probably be considered a typical pattern with a relatively stable basis level throughout the marketing year with a slight seasonal increase from September through January and a decline through March followed by a slight increase from June through August. Analog G2 shows a general weakening (decline) in the basis level throughout the marketing year. Analog G3 has a relative stable pattern from September through June with a slight increase in the basis level in the final two months of the marketing year. Analogs G4 and G5 are similar in that they show a general weakening of the basis through April with a strengthening through the end of the marketing year. The main difference is that G5 (outlier) shows much more volatile swings in the basis when compared to G4. Generally, a cluster that only contains one observation is considered as a potential ‘outlier’ observation. The 2013/14 marketing year, with its extreme fluctuation in the monthly basis levels, is assigned to a singular analog (G5).

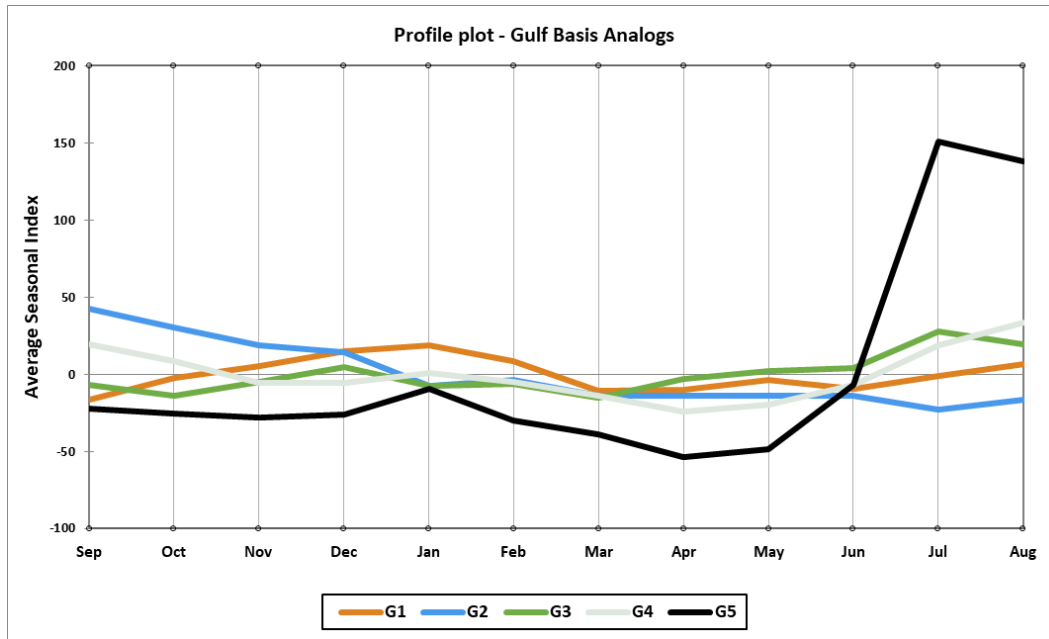


Figure 4



Application of the AHC clustering procedure to the PNW basis data results in the dendrogram shown in Figure 5. The procedure results in four distinct seasonal analogs in the PNW basis (compared to the five for the Gulf) with one single observation analog (P4) composed of the 2014/15 marketing year.

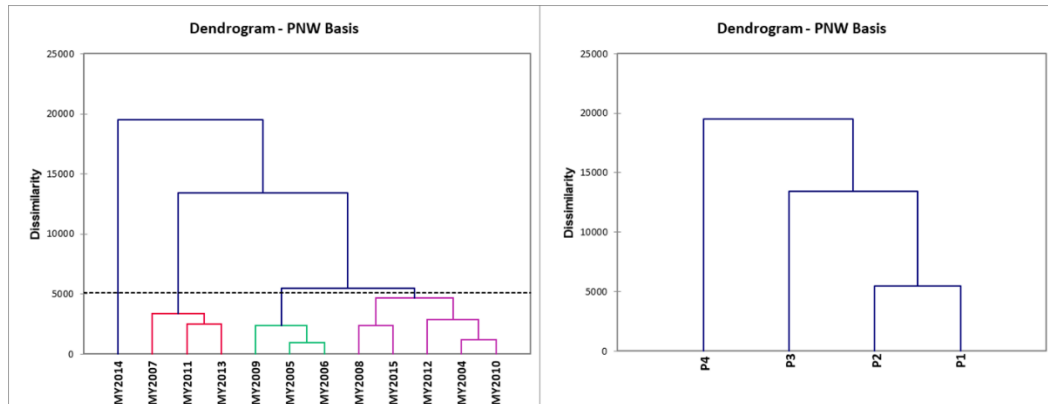


Figure 5

Table 5 provides a summary of the AHC analog groupings for the PNW basis. Analog P1 contains 5 of the marketing years with 3 each in P2 and P3. Analog P4 contains the single 2014/15 marketing year.

Table 5. Summary of AHC Seasonal Analog Groupings for PNW Basis

Analog	P1	P2	P3	P4
Objects	5	3	3	1
Sum of weights	5	3	3	1
Within-class variance	2777.65	1675.66	2926.84	0.00
Minimum distance to centroid	36.89	26.07	42.19	0.00
Average distance to centroid	46.71	32.93	44.11	0.00
Maximum distance to centroid	56.21	40.01	47.32	0.00
	MY2004	MY2005	MY2007	MY2014
	MY2008	MY2006	MY2011	
	MY2010	MY2009	MY2013	
	MY2012			
	MY2015			

The profiles of the four PNW seasonal analogs are shown in Figure 6. Analog P1 is similar to G1 in that it shows a strengthening basis through the first four months of the marketing year (through January) with a weakening through June and a strengthening pattern through the final two months of the marketing year. This behavior is probably most typical. Analog P2 shows a general weakening of the basis throughout the marketing year with a slight uptick in the final month. Analog P3 does not correspond with any of the Gulf patterns, showing a substantial strengthening through the first three months followed by a relatively stable pattern before strengthening in the final three months. Analog G4, as with P5, shows a highly variable pattern; however, it has a sharp weakening in the first two months with an uneven weakening pattern through the end of the marketing year.

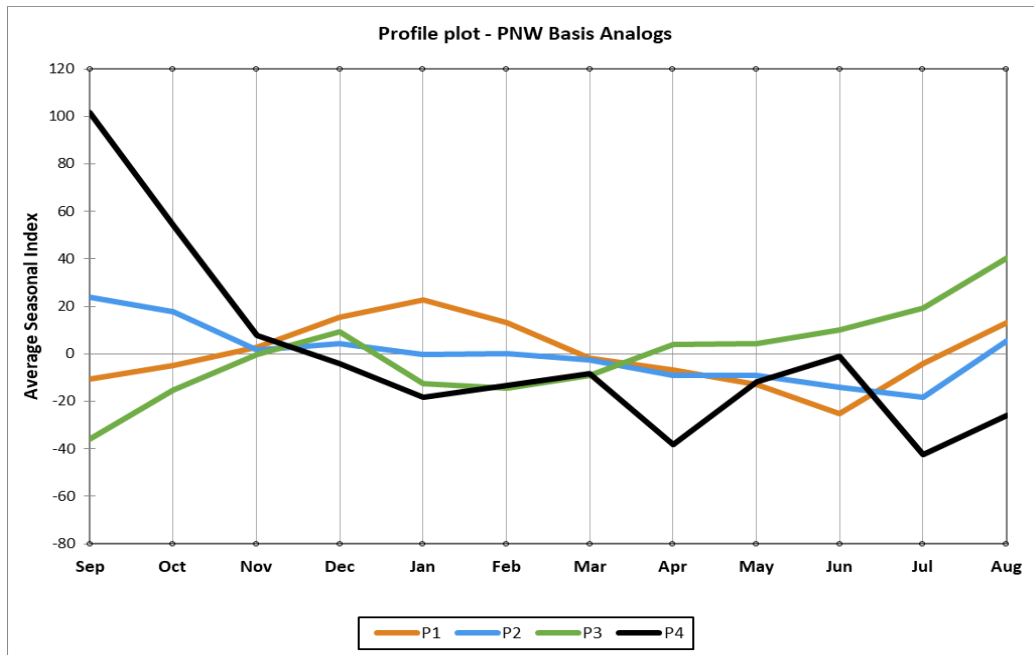


Figure 6

### ***Statistical Characterization of the Basis Seasonal Analogs for Gulf and PNW***

While the marketing year basis levels were highly correlated, the seasonal analogs between the Gulf and PNW do not appear to have much correlation. First, the polychoric correlation coefficient between the analog categories was equal to -0.0963 – an insignificant number. Second, a contingency table was constructed between the two categories and a Chi-square test was conducted on the table to test for dependence between the two analogs. The Chi-Square test value of 12.53 was short of the critical value of 18.55 at the 90% confidence level.

The Lebart z-scores for the Gulf and PNW basis analogs are shown in Table 6. The sign of the z-score indicates whether the analog mean was less than (negative) or greater than (positive) the overall mean value for the explanatory variable. The test is two-tailed; therefore, those z-scores exceeding 1.64 in absolute value are significant at the 90 percent confidence level. Those absolute scores exceeding 1.96 are significant at the 95 percent level and those exceeding 2.56 are significant at the 99 percent level.

Table 6. Variable Characterization Test Z-Scores (2-tailed) Shaded by Sign and Significance

Variable	Gulf Analogs					PNW Analogs			
	G1	G2	G3	G4	G5	P1	P2	P3	P4
Futures-NB	-1.3221	-0.9670	<b>1.9841</b>	-0.4625	0.9914	0.1741	<b>-1.8721</b>	<b>1.8139</b>	-0.2193
FutSprd1	0.8537	1.0025	-0.3557	-0.1158	<b>-1.9506</b>	-0.7959	1.4171	-0.6797	0.2643
FutSprd2	0.7560	1.1533	-0.5989	0.3091	<b>-2.2855</b>	-0.3084	1.4799	<b>-1.4333</b>	0.4771
Basis-Brz	-0.5218	-0.5276	-1.0617	1.0609	1.5300	0.3499	-0.8446	0.0592	0.6063
Rail-Gulf	-1.4324	-0.1886	0.6948	0.1008	1.2520	0.0805	<b>-1.7061</b>	1.0604	0.8679
Rail-Sprd	-1.6278	0.6239	0.7154	-0.5322	1.4220	-1.0496	-1.2067	1.5090	1.3987
Barge-Spot	-0.6086	0.4473	1.0774	-1.5984	1.1665	-1.3040	-0.0137	0.8774	0.9729
Barge-3M	-1.3548	0.8063	1.1246	-1.0208	0.8728	<b>-1.6655</b>	0.2029	1.1082	0.9167
Ocean-PNW	-0.3136	-0.0204	1.4044	-0.8681	-0.3214	-1.2805	0.5331	1.4837	-0.8757
Ocean-Sprd	-0.6890	-1.0079	<b>1.9094</b>	-0.4298	0.1204	-0.6827	-0.6685	<b>1.8968</b>	-0.7067
Gulf-InPort	<b>-2.3784</b>	0.1052	-0.1419	1.2221	<b>1.8919</b>	-0.2917	-0.3449	0.2382	0.6874
PNW-InPort	-1.2923	-0.7809	0.5807	0.0593	<b>2.0750</b>	0.2083	-1.2355	0.7561	0.3794
Cars-Late	0.2039	0.4848	-0.2392	<b>-1.7558</b>	<b>2.1525</b>	-1.4814	-0.0275	1.0196	1.0879
DCV	-0.5291	0.1458	-0.8116	-0.7282	<b>3.0448</b>	-1.2155	-0.2386	1.2172	0.6350
Export-Gulf	-1.4294	0.5159	-0.5318	1.1246	0.6153	0.6755	-1.3337	-0.5237	<b>1.7052</b>
Export-PNW	<b>-1.7890</b>	0.5412	-0.2804	1.3517	0.3946	-0.0662	-1.2701	0.2667	<b>1.6900</b>
Export-Out	<b>-1.8365</b>	0.2369	0.4077	0.3779	1.3268	0.0012	<b>-1.4439</b>	0.5798	1.3516
FarmDel-Q1	-1.6246	1.4614	1.5198	-1.4150	0.4105	-0.9666	-0.7861	0.4717	<b>2.2169</b>
FarmDel-Q2	-0.9311	0.6869	1.0198	-1.4632	1.2273	-1.1553	-0.4877	1.0198	1.2273
FarmDel-Q3	-0.2351	0.4943	-0.1008	-0.7724	1.0698	-1.0127	-0.3694	1.2426	0.4385
Crush	1.2143	-1.2426	0.8878	-0.6648	-0.5763	-0.3471	0.2540	1.0485	-1.4214
SU-Ratio	1.0078	1.4966	-0.6127	-1.0564	-0.9820	-0.8911	<b>2.3223</b>	-1.0301	-0.4349
MealP	-1.5326	-0.6231	1.1011	-0.1038	<b>1.6790</b>	0.0832	<b>-1.9267</b>	1.5621	0.4230
OilP	-0.8512	-0.8653	<b>2.4217</b>	-0.8724	0.0732	-0.1028	-1.3321	<b>1.8039</b>	-0.5559
World-SU	-0.5925	<b>1.6760</b>	0.0257	-0.5242	-0.5506	-0.5224	1.6038	-1.4415	0.6776
SA-Prod	<b>-1.8621</b>	0.9109	0.2079	0.2660	0.9465	-0.1294	-0.8381	-0.1138	<b>1.7222</b>
China-Import	-1.3782	0.1778	-0.1309	0.7074	1.0163	0.1757	-1.5809	0.4606	1.4419

\*Values in **bold** are significant at 90% confidence level, **bold italics** at 95% confidence level.

For the Gulf basis, the results indicate that analog G1 is mostly associated with marketing years with lower than average weekly ships in port at the Gulf, lower than average South American soybean production, lower than average weekly outstanding exports, and lower than average weekly export inspections in the PNW. Analog G2 is characterized by years with a higher than average world soybean stocks-use ratio. Analog G3 is characterized by higher than average domestic soybean oil prices, higher than average nearby futures prices, and higher than average PNW ocean freight costs (relative to the Gulf). Analog G4 is characterized by lower than average weekly railcars placed late. Analog G5 is characterized by higher than average secondary railcar values, lower than average second nearby futures carry spreads, a higher than average weekly number of railcars placed late, a higher than average weekly number of ships in port at the PNW, a lower than average nearby futures carry spread, a higher than average weekly number of ships in port at the Gulf, and a higher than average domestic soybean meal price. All of these significant at the 90% confidence level or higher.

For the PNW basis, analog P1 is characterized by a lower than average weekly forward barge rate. Analog P2 is characterized by a higher than average domestic soybean stocks-use ratio, a lower than average domestic soybean meal price, a lower than average nearby futures price, and a lower than average weekly rail cost from Freemont, NE to the Gulf. Analog P3 is characterized by a higher than average weekly ocean freight cost for the PNW relative to the

Gulf, a higher than average nearby soybean futures price, and a higher than average domestic soybean oil price. Analog P4 is characterized by a higher than average percentage of farmer deliveries in Q1, a higher than average South American soybean production, and higher than average export inspections out of both the Gulf and PNW. All of these significant at the 90% confidence level or higher.

For all of these seasonal analogs, the average Brazilian export basis was not statistically a significant factor nor was the overall level of Chinese imports. This is unlike the results of the average basis analysis. However, it appears that logistical costs (barge and ocean rates in addition to rail costs) and conditions (cars placed late, secondary railcar market values, and pace of farmer marketings) play a more important role in determining the seasonality basis as opposed to its overall average level.

## **Summary and Conclusions**

There is substantial variability in the basis for crops and oilseeds at the export market. While this has been an issue for some time, the basis level and variability has increased over time. Other studies have mostly analyzed basis at the futures delivery market, or at crop origins. But, factors impacting the basis in these locations differ substantially from those impacting export basis values. In fact, some studies use the export basis as an explanatory value in the analysis of origin basis. The export basis is highly variable, both inter year and intra year and potentially explained by numerous factors including selected world supply and demand conditions, the rate of importing by major buyers, international competing basis values, as well as intramarket spreads and shipping costs. Export basis are also highly seasonal, but the seasonal behavior varies across marketing years. Commodity analysis refer to this as analogue years and variations in seasonal behavior is an important feature in understanding markets.

The purpose of this study is to examine the impact of market supply/demand and logistical variables upon both the average level and seasonality of U.S. export basis values for soybeans for the period 2004/05 through 2015/16. We developed and estimated models using statistical regression technique, called partial least squares (PLS) is utilized. To explain seasonality, this study utilizes agglomerative hierarchical clustering (AHC) to cluster similar marketing years by seasonal pattern called seasonal analogs. These seasonal analogs were then related to explanatory variables using a version of PLS that includes discriminant analysis. Additionally, a statistical test (variables characterization test) compares the means of a subset and its parent set to explain the impact of the explanatory variables.

The results indicate that the marketing year average basis level for the Gulf and PNW markets is primarily influenced by international and domestic competitive pressures. The most significant variable impacting both the Gulf and PNW export basis values is the Brazilian export basis which the U.S. export markets respond in order to be competitive in the international export market (particularly to China). The positive sign of the coefficient (in both markets) indicates that the U.S. export basis must partially match changes in the Brazilian export basis.

The U.S. marketing year average soybean export basis levels are also significantly impacted by the level of soybean imports by China – even after adjusting for the upward trend in both variables. It is important that China is by far the largest soybean buyer from both port areas. While the dominance of China to the PNW market is well known, the results show that China is also of paramount importance to the Gulf export basis values with significant and positive coefficient values in both markets.

The futures carry spreads (deferred minus nearby CME futures prices) are also highly influential on the export basis with a highly significant coefficient values with inverse signs in both equations. The same effect is also indicated by the significant and positive sign on the coefficients for the domestic soybean meal price. A higher domestic soybean meal price is an indicator of strengthened domestic demand which requires stronger soybean export basis values to maintain competitive parity with domestic markets.

The results indicate that the marketing year average export basis levels are more sensitive to domestic rail transportation costs (tariff plus fuel surcharge) when compared to barge and ocean freight. Both variables had the expected signs (positive) which reflect how the basis responds to the higher transportation costs in order to encourage shipments to the export ports.

Of significant, but lesser importance, is the negative influence of the domestic stocks-use ratio and the positive influence of the current nearby soybean futures price level upon both the Gulf and PNW export basis values. Higher stocks-use ratio values are indicative of a high level of domestic supply relative to total (domestic and export) demand during the marketing year. These signs are as expected and fit with the previous results which indicate that export basis levels adjust to maintain competitiveness with the domestic market and assure a critical flow of soybeans into the export channels to meet demand.

The AHC clustering procedure identified a total of 5 and 4 distinct seasonal basis patterns for the Gulf and PNW respectively. In each market, there is a distinct outlier analog containing just one marketing year (2013/14 for Gulf, 2014/15 for the PNW) and these particular analogs were characterized by extreme seasonal swings in the basis. Unlike the marketing year average basis level, which was highly correlated (93 percent) between both ports, the application of polychoric correlation and Chi-squared tests indicated that the correlation of seasonal analogs between the two ports was essentially nil. This indicates that seasonal patterns tend to be unique to each port while the overall average level of the basis is not.

Application of the Lebart z-score test indicated that the seasonal analogs had varying and diverse sets of explanatory variables – however, these were mostly related to variables characterizing the level of export activity, farmer marketings, and logistical situations. In particular, the singleton outlier analogs (G5 and P4) for both markets, which exhibited high basis variability, were dominated by the presence of logistical conditions (i.e., high number of late railcar placements, high secondary railcar values) in the face of very high levels of demand and commodity flows (higher levels of farmer marketings earlier in the marketing year, high numbers of ships in port and export inspections, low domestic stocks-use ratios).

The primary contributions of this study can be broken down into three distinct observations. First, the marketing year average level of export basis in the two primary U.S. export markets (Gulf and PNW) is primarily driven by the need to maintain competitive parity – both with the international (Brazil) and domestic markets. U.S. export basis values adjust with changes in competitor's export basis values (in U.S. dollars per bushel) as expected. Likewise, export basis values adjust accordingly with domestic demand (and prices) in order to assure that an adequate supply of soybeans move to the export channels to meet anticipated export demands.

Second, seasonality in the Gulf and PNW export basis values is not consistent across marketing years. As a result, it is necessary to realize that different market conditions result in different seasonal basis patterns called seasonal analogs in the trade. In this study, over the 12 marketing year period analyzed (2004/05 through 2015/16), a total of five and four unique seasonal analog patterns were derived using statistical clustering applications. For each market, one of these patterns was a unique year (outlier) characterized by extremely high basis volatility. The high number of analogs (5 for Gulf and 4 for PNW) over the short time frame considered (12 marketing years) indicates that seasonality of the basis is highly unstable from year-to-year.

Third, unlike the average marketing year basis level, the seasonal analogs patterns are primarily driven by three categories mostly unique to each export market: (1) the level of export activity at the particular port, (2) the pace of farmer marketings throughout the marketing year, and (3) the logistical conditions (lateness of railcar placement, cost of secondary railcars, barge and ocean freight rates) present in the marketing year along with transportation cost differentials (between the two ports and primarily barge and rail). In particular, the extreme outlier analogs were both characterized by strong nearby demand and complicating logistical conditions.

These results suggest a number implications. First, analysis of export basis values is far more complicated than analysis of basis at futures delivery or crop origin markets. The export basis is also impacted by numerous variables including world supply and demand, basis at competing markets, logistics costs and performance, intermonth prices spreads, farmer deliveries, etc.

Second, our study analyzes the soybean export basis from the United States, for which the basis at a competing market, Brazil, is of great importance. Ultimately, the competing basis value are connected through spatial competition. One would expect that similar phenomena would be apparent in other crops including corn where Ukraine and Argentine export basis would be important, and wheat, where export basis from Canada, Australia, and Black Sea would be important. All of these would compound analysis of crop origin basis.

A third implication is the role of logistics and shipping costs. It is commonly recognized that shipping costs and functions impact the export basis. These results indicate that while logistics is important in varying ways, other variables are of greater importance. Here, that would include basis at competing export markets, and import demand from the dominant buyer, China. Following these factors, shipping costs and logistics are important in a logical way and include impacts of farmer deliveries, intermonth futures prices spreads (impacting storage decisions) as well as barge rates and rail performance (rail cars late) and ancillary shipping costs (secondary

rail market values). Further, that seasonal behavior of the basis is not the same among crop marketing years, would only compound rail shipping and logistics management decisions.

Finally, it is important to note that these results indicate the current volatility of export basis, the likely escalation in future volatility, and the heterogeneous seasonal behavior of export basis values. This is in a world whereby United States pricing mechanisms, which in this case are concentrated at Chicago and built around instruments really focused on domestic demand, supply and logistics in addition to a common exchange rate. Traditionally, the stability and predictability of the basis is what promotes hedging using these instruments. However, with the growth in production regions located outside of the United States, the implications of basis volatility has become more apparent. There is no doubt this has been a challenge of U.S. futures markets and is probably what has prompted emergence of a number of non-traditional mechanisms for risk management. This would include development of MATIF. More recently, there has been new contracts focused on off-shore locations and using financial swaps by both PLATTS and McDonald Pelz. Indeed, the CME has begun trading using PLATTS prices in several new contracts. These are in addition and/or complementary to more complex OTC instruments. To date new market mechanisms, have been developed for Black Sea wheat and corn. In addition, Australia wheat and Brazil export soybeans are currently under investigation. If and as these become adopted in the commercial market place, it will provide challenges and opportunities for the entire trading industry and having an understanding of factors impacting export basis will become critical.

## References

- Abdi, H. 2007. "Partial Least Square Regression PLS-Regression." In N. Salkind, ed. *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage, p. 13.
- Adjemian, M., P. Garcia, S.H. Irwin, and A. Smith. 2013. "Non-Convergence in Domestic Commodity Futures Markets: Causes, Consequences, and Remedies." No. 115, Available at: <http://www.ssrn.com/abstract=2323646> [Accessed January 22, 2019].
- Alexander, R., Z. Zhao, E. Szekely, and D. Giannakis. 2017. "Kernel Analog Forecasting of Tropical Intraseasonal Oscillations." *Journal of Atmospheric Sciences* 74(4):1321–1342.
- Bekkerman, A., G.W. Brester, and M. Taylor. 2016. "Forecasting a Moving Target: The Roles of Quality and Timing for Determining Northern U.S. Wheat Basis." *Journal of Agricultural and Resource Economics* 41(1):25–41.
- Bullock, D.W. 2004. "Seasonal Forecasting of Minnesota HRS Wheat Price and Basis." Minnesota Department of Agriculture.
- Comeau, D., D. Giannakis, Z. Zhao, and A.J. Majda. 2017. "Predicting regional and pan-Arctic sea ice anomalies with kernel analog forecasting." *arXiv preprint arXiv:1705.05228*.
- Coyle, T. 2016. "CME Delivery Economics."
- Dhuyvetter, K.C., and T.L. Kastens. 1998. "Forecasting Crop Basis: Practical Alternatives." In *Proceedings of the NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management*. 1998 NCR-134 Conference. Chicago, IL.
- Dhuyvetter, K.C., K. Swanser, T. Kastens, J.R. Mintert, and B. Crosby. 2008. "Improving Feed Cattle Basis Forecasts." In *Proceedings of the Western Agricultural Economics Association Annual Meeting*. 2008 Western Agricultural Economics Association Annual Meeting. Big Sky, MT.
- Djalalova, I., L. Delle Monache, and J. Wilczak. 2015. "PM2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model." *Atmospheric Environment* 119:431–442.
- Elmore, R.W., and S.E. Taylor. 2013. "Analog years for weather forecasting and correlating corn planting dates with yield in Iowa." Integrated Crop Management News No. 82, Iowa State University. Available at: <http://lib.dr.iastate.edu/cropnews/82>.
- Esbensen, K.H., and B. Swarbrick. 2018. *Multivariate Analysis* 6th ed. Oslo: CAMO Software AS.
- F. Hair Jr, J., M. Sarstedt, L. Hopkins, and V. G. Kuppelwieser. 2014. "Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research." *European Business Review* 26(2):106–121.



- Flaskerud, G., and D. Johnson. 2000. "Seasonal Price Patterns for Crops." No. EB-61, North Dakota State University.
- Garner, C. 2015. "Gaming the Seasonality in Commodities: What you should be looking for in coming months." *Real Money*. Available at: <https://realmoney.thestreet.com/articles/01/30/2015/gaming-seasonality-commodities> [Accessed January 23, 2019].
- Gray, R.W., and A.E. Peck. 1981. "The Chicago wheat futures market: Recent problems in historical perspective." *Food Research Institute Studies* 18(1):89–115.
- Hansen, J.W., A. Potgieter, and M.K. Tippet. 2004. "Using a general circulation model to forecast regional wheat yields in northeast Australia." *Agricultural and Forest Meteorology* 127(1–2):77–92.
- Hart, C., and F. Olson. 2017. "Analysis of Grain Basis Behavior During Transportation Disruptions and Development of Weekly Grain Basis Indicators for the USDA Grain Transportation Report." No. 82, Center for Agricultural and Rural Development (CARD) at Iowa State University. Available at: [https://lib.dr.iastate.edu/card\\_staffreports/82](https://lib.dr.iastate.edu/card_staffreports/82).
- 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.
- Hieronymus, T. 1994a. "Affidavit of Professor Thomas A. Hieronymus re: CFTC versus Ferruzzi."
- Hieronymus, T. 1994b. "Expert Report of Professor Thomas Hieronymus re: CFTC versus Ferruzzi."
- Hieronymus, T.A. 1971. *Economics of Futures Trading For Commercial and Personal Profit*. New York: Commodity Research Bureau.
- Irwin, S., and D. Good. 2016. "Forming Expectations for the 2016 US Average Soybean Yield: What About El Niño?" *farmdoc daily* No. (6):46, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. Available at: <http://farmdocdaily.illinois.edu/2016/03/expectations-for-2016-us-average-soybean-yield.html>.
- Johansson, R., E. Luebehusen, B. Morris, H. Shannon, and S. Meyer. 2015. "Monitoring the impacts of weather and climate extremes on global agricultural production." *Weather and Climate Extremes* 10:65–71.
- Kluis, A. 2017. "Corn, Soybean Seasonal Patterns are Changing." *Successful Farming*. Available at: <https://www.agriculture.com/markets/analysis/crops/corn-soybean-seasonal-patterns-are-changing> [Accessed January 23, 2019].
- Kolb, R.W., and J.A. Overdahl. 2007. *Futures, Options, and Swaps* 5th ed. New York: Wiley-Blackwell.

- Kolton, A.D., R.A. Gamboa, and D.S. Chimenti. 1996. "System for forming queries to a commodities trading database using analog indicators."
- Kub, E. 2015. "Insufficient Freight: An Assessment of US Transportation Infrastructure and Its Effects on the Grain Industry."
- Kuhn, M., and K. Johnson. 2013. *Applied Predictive Modeling*. New York: Springer.
- Lahmiri, S., G.S. Uddin, and S. Bekiros. 2017. "Clustering of short and long-term co-movements in international financial and commodity markets in wavelet domain." *Physica A: Statistical Mechanics and its Applications* 486:947–955.
- Lara-Chavez, A., and C. Alexander. 2006. "The Effects of Hurricane Katrina on Corn, Wheat, and Soybean Futures Basis." In *Proceedings of the NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management*. 2006 NCR-134 Conference. St. Louis, MO.
- Lebart, L., A. Morineau, and M. Piron. 2000. *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Leuthold, R., J. Junkus, and J. Cordier. 1989. *The Theory and Practice of Futures Markets*. Lexington, MA: Lexington Books.
- Levine, M. 2015. "Oreos, vomitoxin and the price of wheat." *Crains Chicago Business. com*.
- Manfredo, M.R., and D.R. Sanders. 2006. "Is the Local Basis Really Local?" In *Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. 2006 NCCC-134 Conference. St. Louis, MO.
- Martens, H., and T. Naes. 1989. *Multivariate Calibration*. London: Wiley.
- Massy, W. 1965. "Principal Components Regression in Exploratory Statistical Research." *Journal of the American Statistical Association* 60:234–246.
- Menzie, K. 2007. "Methods of Evaluating Agrometeorological Risks and Uncertainties for Estimating Global Agricultural Supply and Demand." In M. V. K. Sivakumar and R. P. Motha, eds. *Managing Weather and Climate Risks in Agriculture*. Springer, pp. 125–140.
- Olson, F. 2014. "Effects of 2013/14 Rail Transportation Problems on North Dakota Farm Income." North Dakota State University.
- Onel, G., and B. Karali. 2014. "Relative Performance of Semi-Parametric Nonlinear Models in Forecasting Basis." In *Proceedings of the American Agricultural Economics Association Annual Meeting*. 2014 AAEA Annual Meeting. Minneapolis, MN.
- Parcell, J.L. 2000. "The Impact of the LDP on Corn and Soybean Basis in Missouri." In *Proceedings of the NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management*. 2000 NCR-134 Conference. Chicago, IL.

- Peck, A.E., and J.C. Williams. 1991. "Deliveries on the Chicago Board of Trade Wheat, Corn, and Soybean Futures Contracts, 1964/65 to 1988/89." In W. P. Falcon, W. O. Jones, C. H. Gotsch, and L. W. Perry, eds. *Food Research Institute Studies*. Palo Alto, CA: Food Research Institute, Stanford University, pp. 129–225.
- Phillips, G.M., W.P. Jennings, M.C. Findlay III, S.A. Klein, and M.E. Rice. 2004. "Combination forecasting using clusterization."
- Pirrong, C. 2004. "Detecting Manipulation in Futures Markets: The Ferruzzi Soybean Episode." *American Law and Economics Association* 6(1):28–71.
- 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.
- Seamon, F. 2010. "Understanding Wheat Futures Convergence." CME Group. Available at: [www.cmegroup.com/trading/agricultural/files/AC-422\\_WheatResearchPaper\\_FINAL\\_SR.PDF](http://www.cmegroup.com/trading/agricultural/files/AC-422_WheatResearchPaper_FINAL_SR.PDF).
- Seamon, V.F., K.H. Kahl, and C.E. Curtis, Jr. 2001. "Regional and Seasonal Differences in the Cotton Basis." *Journal of Agribusiness* 19(2):147–161.
- Shannon, C.E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27(3):379–423.
- 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.
- 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.
- 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.
- Tonsor, G.T., K.C. Dhuyvetter, and J.R. Mintert. 2004. "Improving Cattle Basis Forecasting." *Journal of Agricultural and Resource Economics* 29(2):228–241.
- USDA Office of the Chief Economist and Agricultural Marketing Service. 2014. "Rail Service Challenges in the Upper Midwest: Implications for Agricultural Sectors – Preliminary Analysis of the 2013-14 Situation." U.S. Department of Agriculture, Office of the Chief Economist and the Agricultural Marketing Service.
- Usset, E. 2014. "Minnesota Basis Analysis: Preliminary Report for the Minnesota Department of Agriculture." Center for Farm Financial Management, University of Minnesota.

- Villegas, L. 2016. "No Train no Grain: The impact of increased demand for rail services by the energy sector on wheat prices." *Indian Journal of Economics and Development* 4(4).
- Wanat, S., S. Śmiech, and M. Papież. 2016. "In Search of Hedges and Safe Havens in Global Financial Markets." *Statistics in Transition new series* 3(17):557–574.
- Ward, J.H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58(301):236–244.
- Welch, J.M., V. Mkrtchyan, and G.J. Power. 2009. "Predicting the Corn Basis in the Texas Triangle Area." *Journal of Agribusiness* 27(1/2):49–63.
- Wilson, W.W., and B. Dahl. 2011. "Grain pricing and transportation: dynamics and changes in markets." *Agribusiness* 27(4):420–434.
- Wold, H. 1966. "Estimation of principal components and related models by iterative least squares." In P. R. Krishnaiah, ed. *Multivariate Analysis*. New York: Academic Press, pp. 391–420.
- Wold, H. 1973. "Non-linear Iterative Partial Least Squares (NIPALS) modelling: Some current developments." In P. R. Krishnaiah, ed. *Multivariate Analysis III*. New York: Academic Press, pp. 383–407.
- Wold, S. 2001. "Personal memories of the early PLS development." *Chemometrics and Intelligent Laboratory Systems* 58:83–84.
- Wold, S., M. Sjostrom, and L. Eriksson eds. 1993. *PLS-partial least squares projections to latent structures*. Leiden: ESCOM Science Publishers.
- Working, H. 1953. "Hedging Reconsidered." *Journal of Farm Economics* 35(4):544–561.
- Working, H. 1949. "The theory of price of storage." *The American Economic Review* 39(6):1254–1262.
- Zhang, R., and J. Houston. 2005. "Effects of Price Volatility and Surging South American Soybean Production on Short-Run Basis Dynamics." In *Proceedings of the NCR-134 Conference on Applied Commodity Analysis, Forecasting, and Market Risk Management*. 2005 NCR-134 Conference. St. Louis, MO.