5-2017

A Tournament Approach to Price Discovery in the US Cattle Market

Jeffrey Wright

Utah State University

A TOURNAMENT APPROACH TO PRICE DISCOVERY IN THE U.S. CATTLE MARKET

by

Jeffrey Wright

A thesis submitted in the partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Economics and Statistics

Approved:

Man-Keun Kim Major Professor	Dillon M. Feuz Committee Member
Hernan Tejeda Committee Member	Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2017

ABSTRACT

A Tournament Approach to Price Discovery in the US Cattle Market

by

Jeffrey Wright, Master of Science

Utah State University, 2017

Major Professor: Dr. Man-Keun Kim
Department: Applied Economics

Price dynamics among regional cattle markets and futures markets can provide useful information for delineating a relevant fed cattle procurement market. The dominant markets where information is discovered first may play the role of price leader providing substantial market information to other markets, or price followers, which may have insufficient activity to generate much new information. Locating the price discovery center or market, and estimating price interactions among the regional fed cattle markets and also among feeder cattle markets can help define a relevant fed cattle procurement market. This study examines the development of these cattle markets, fed and feeder, into distinct submarkets. It then investigates the dynamic relationships among the US cattle markets across regions, cattle types, and cash/futures markets. The comparison of many markets by using an error correction model is accomplished with the introduction of a tournament approach, and refined by combining this approach with a hierarchical cluster analysis. Application of these techniques allows us to conclude that the appropriate price for the U.S. cattle markets is discovered in the futures markets, feeder cattle futures and fed futures.

(51 pages)

PUBLIC ABSTRACT

A Tournament Approach to Price Discovery in the US Cattle Market

Jeffrey Wright

Cattle price discovery is a process of determining the price in the market through the interactions of cattle buyers (packers) and sellers (ranchers). Locating the price discovery center or market, and estimating price interactions among the regional fed cattle markets and also among feeder cattle markets can help define a relevant fed cattle procurement market. This research identifies that the U.S. cattle markets is discovered in the futures markets, feeder cattle futures and fed futures.

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Man-Keun Kim, for his immense support, guidance, and encouragement throughout my work without which I could never have completed this work. I also want to thank my committee members, Dr. Dillon Feuz and Dr. Hernan Tejeda, for their support, valuable suggestion, and assistance in the entire process. Lastly, I want to express gratitude to my wife Nadia for her immense help and support.

Jeffrey Wright

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Price dynamics among regional cattle markets and futures markets can provide useful information for delineating a relevant fed cattle procurement market. The dominant markets where information is first discovered may play the role of price leader providing significant market information to other markets, price followers, which may have insufficient activity to generate new information (Schroeder, 1996). Locating the price discovery center or market, and estimating price interactions among the regional fed and feeder cattle markets, can help define a relevant fed cattle procurement market.

The term price discovery refers to a process whereby the relative contributions of interrelated submarkets to the overall market price can be determined. The submarket with the larger contribution is called the “price discovery (market).” Joseph, Garcia, and Peterson (2013) emphasized that “effective price discovery is critical as it facilitates pricing quantity and quality of a commodity at a specified time and place” (page 1).

Several studies have investigated price leadership and delineation of the relevant geographic market for fed cattle. Koontz, Garcia, and Hudson (1990), using weekly prices from 1973 through 1984 and the Granger causality analysis, found that the Nebraska direct market responded fastest to new information. Through the vector autoregression (VAR) of weekly regional fed cattle prices from 1976 through 1987, Schroeder and Goodwin (1990) found that Iowa/Southern Minnesota and Eastern Nebraska tended to be the leading price discovery regions, with western Kansas becoming more dominant over the time period.

Schroeder (1996) used plant-level transaction prices (from March 23, 1992 to April 3, 1993) and applied VAR models and Granger causality analysis, finding that slaughter plants in Kansas and Nebraska tended to be price discovery leaders. Plants in other states reacted most quickly to price changes at the Nebraska plants. Lee and Kim (2007) investigated the dynamic relationship among fed cattle markets applying an Error Correction Model (ECM) to weekly slaughter steer prices from February 25, 2002 to July 3, 2006 and found that the Kansas market is a dominant price leader - result contrary to previous studies.

Regarding studies with cattle futures prices, Oellermann, Brorsen, and Farris (1989) found that futures prices of feeder cattle explained cash prices but not the reverse. Joseph, Garcia, and Peterson (2013) studied price discovery in the U.S. fed cattle market, examining the interaction among weekly live cattle futures, negotiated cash fed cattle, and boxed beef cutout prices. Extensive testing and innovation accounting based on directed acyclic graphs of error-correction residuals indicates that the futures price continues as the dominant source of information in the fed cattle market. While the cash cattle price has a strong predictive influence on the boxed beef price, the boxed beef price plays only a marginal role in price discovery.

Although causality testing can be informative, these tests focus on the impact of lagged prices on current prices (Arnade and Hoffman, 2015). Instead in this study, relative adjustment rates derived from the estimated (bivariate) error correction model (ECM) are used to estimate price discovery weights, which was developed by Gonzalo and Granger (1995) and expanded in Theissen (2002). The technique, explained in detail in the next section, is a function of the relative ratio of the speed of adjustments in an ECM. This method is used by Schwarz and Szakmary (1994), Foster (1996), Theissen (2002), Eun and Sabherwal (2003), Thurlin (2009), Figuerola-Ferretti and Gonzalo (2010), Plato and Hoffman (2011), Kim (2011), and Arnade and Hoffman (2015) to identify price discovery among related markets.

This technique is applicable only for the bivariate case, i.e., one cointegration vector, and not applicable for the multivariable case that may have multiple cointegrating vectors. This study suggests a novel way to overcome this problem by using the cluster analysis and tournament approach. This technique allows us to compare many cattle market prices across regions, cattle type, and cash/futures markets

This study investigates the dynamic relationships among the US cattle markets across regions, cattle types, and cash/futures markets. This study attempts to determine the reference US cattle price. This study is structured as follows. Chapter 1 provides a general introduction to the price discovery process and discusses the research objectives. Chapter 2 provides the overview of the U.S. cattle industry and chapter 3 explores the concept of price discovery and introduces cluster analysis. Chapter 4 explains data used and chapter 5 presents results and discussion. Chapter 6 draws conclusions of the study and outlines future studies.

OVERVIEW OF THE U.S. CATTLE INDUSTRY

With the largest fed-cattle industry in the world, the U.S, is also the world's largest producer of beef, primarily high-quality, grain-fed beef for domestic and export use (USDA, 2016). Cattle production is one of the most important industries in the U.S. accounting for $78.2 billion in cash receipts during 2015 (USDA, 2016). The U.S. cattle and calves inventory peaked in 1975 with 132 million head. With a few cattle cycles[1] as shown in Figure 1, cattle and calves inventory in 2016 was 92 million head, 3 percent above the 89 million head in 2015. The 92.0 million head in 2016 is the largest number of cattle and calves since 2011 (Figure 1). The U.S. cattle industry is roughly divided into two production sectors: cow-calf operations and cattle feeding.

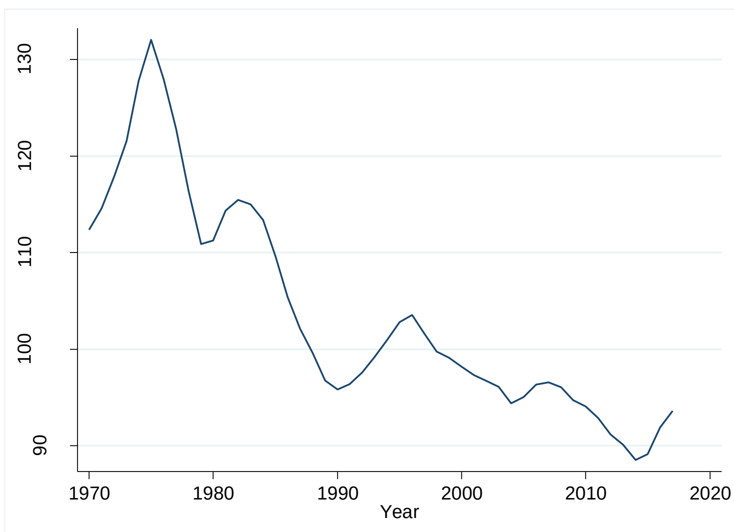

Figure 1. Cattle inventory (Jan 1st) in million head

Source: USDA NASS

[1] Womach (2005, page 41) defines the cattle cycles as "…the approximately 10-year period in which the number of U.S. beef cattle is alternatively expanded and reduced over several consecutive years in response to perceived changes in profitability by producers. Generally, low prices occur when cattle numbers (or beef supplies) are high, precipitating several years of herd liquidation. As cattle numbers decline, prices gradually begin to rise, causing producers to begin adding cattle to their herds."

Cow-calf Operations: Feeder Cattle

USDA (2017) summarizes cow-calf operations in the U.S. such that "Cow-calf operations are located throughout the U.S., typically on land not suited or needed for crop production. These operations depend on range and pasture forage conditions, which in turn depend on variations in the average rainfall and temperature for the area. Beef cows harvest forage from grasslands to maintain themselves and raise calf with very little, if any, grain input. The cow is maintained on pasture year round, same as the calf until it is weaned. If additional forage is available at weaning, some calves may be retained for additional grazing and growth until the following spring when they are sold. The average beef cow herd size is 40 head, but operations with 100 or more beef cows compose nine percent of all beef operations and 51 percent of the beef cow inventory. Operations with 40 or fewer head are largely part of multi-enterprises, or are supplemental to off-farm employment" (USDA, 2017, https://www.ers.usda.gov/topics/animal-products/cattle-beef/background/).

Different marketing options exist for cow-calf producers. The oldest, and still common, is the weekly livestock auction. These auctions can be found throughout the country, wherever significant cattle production takes place. Government mandated price reporting makes these auctions a valuable resource for economists studying commodities markets. In addition to weekly auctions, breeders can also sell their cattle through graded feeder cattle sales. These sales are usually organized by cattleman's associations, sometimes in conjunction with government entities. Cattle, at these sales, are combined into uniform lots. This uniformity helps minimize transportation and processing costs for buyers, allowing sellers to receive the maximum possible value for their calves. These sales also help small producers connect with large producers who are not interested in buying one or two calves at a time. A third option, which has become more popular as the size of cattle operations has increased is the direct farm sale. A typical size for this type of sale is around

50,000 pounds. This kind of sale is desirable to buyers because there is less chance of disease than there would be if calves came from different sources. Cattle from large suppliers are also more accustomed to feedlot conditions and experience less stress and health problems in their new environment. Younger, lighter feeders tend to sell for more dollars per pound than older, heavier feeders, since the potential for gain is greater with the younger cattle. Breed and coloring have also been shown to affect price, with exotic breeds commanding a premium.

Figure 2 presents a typical feeder cattle price (dollars per hundredweight, Nebraska feeder steer) movement over time.

Cattle Feedlots: Fed Cattle

Once feeder cattle are purchased, they are transported to feed lots where they are fattened on grain until the time of slaughter (cattle feeding). Cattle are typically fed until they reach 1,100 to 1,400 pounds. Cattle feeding is concentrated in the Great Plains, but is also important in parts of

Figure 2. Nebraska feeder heifer weekly price ($/cwt)
Source: Livestock Marketing Information Center: Combined Auction

the Corn Belt, Southwest, and Pacific Northwest. Cattle feedlots produce high-quality beef, grade Select or higher, by feeding grain and other concentrates for about 140 days. Depending on weight at placement, feeding conditions, and desired finish, the feeding period can be from 90 to as long as 300 days. Average gain is 2.5-4 pounds per day on about 6 pounds of dry-weight feed per pound of gain. While most of a calf's nutrient inputs until it is weaned are from grass, feedlot rations are generally 70- to 90-percent grain and protein concentrates.

Feedlots with less than 1,000 head of capacity compose the vast majority of U.S. feedlots, but market a relatively small share of fed cattle. In contrast, lots with 1,000 head or more of capacity compose less than five percent of total feedlots, but market 80- to 90-percent of all fed cattle. Feedlots with 32,000 head or more of capacity market around 40-percent of fed cattle. The industry continues to shift toward a small number of very large specialized feedlots, which are increasingly vertically integrated with the cow-calf and processing sectors to produce high-quality fed beef.

As previously mentioned, purchasers of feeder cattle have several different options to obtain this input. Graded feeder sales are arranged in the spring and fall. Feeder cattle can also be obtained throughout the year at weekly auctions. Feeder cattle and feed are the main inputs required to produce fed cattle. Cost can be minimized by buying preconditioned cattle. These calves have been dehorned, castrated, vaccinated, treated for parasites, and started on a grain based diet. The high degree of variability and risk associated with this type of business led to the adoption of futures contracts within the industry. Crop insurance has also become available to some cattle feeder operations.

Figure 3 presents a typical fed cattle price (Nebraska fed steer)

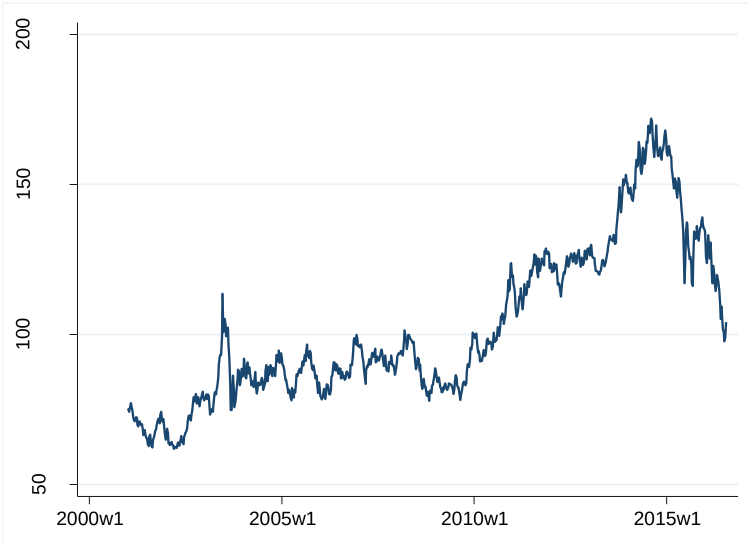

Figure 3. Nebraska fed steer weekly price ($/cwt)

Source: Livestock Marketing Information Center

PRICE DISCOVERY AND CLUSTER ANALYSIS

Price Discovery

Price discovery is about determining which market is more informative. Ward and Schroeder (2002) defines price discovery as "…is the process of buyers and sellers arriving at a transaction price for a given quality and quantity of a product at a given time and place." (page 1) when an asset or similar commodities are traded in different markets such as cattle. A number of different methods have been used to study price discovery. One of the methods is to use bivariate time series analysis with an error correction term and compare speed of adjustment between the two series (Gonzalo and Granger, 2005). A detailed description of these methods is presented in the following sections. The underlying idea behind price discovery, however, can be most easily explained through a graphical example.

Let p_1 represent the cattle price in region 1 (or market 1) and let p_2 represent the price of a similar type of cattle in region 2 (or market 2). Suppose that p_1 is higher than p_2 as shown in Figure 4 and they move together, that is, cointegrated. Suppose that, for some reasons, a shock occurs in market 1, causing p_1 to rapidly increase (first panel, Figure 4). A number of things can occur after this change. Let us assume that market 1 is the price discovery market. Although we will likely see a slight downward correction in p_1 in market 1, we will see a much larger upward change in p_2. These changes are demonstrated in the second panel in Figure 4. This occurs because buyers and sellers in market 2 look to market 1 to ascertain the correct price in market 2. In other words, market 1 market 1 is more informative.

Let us now assume market 2 is the price discovery. In this case, traders in market 2 are much less concerned with market 1. Although there may be a slight increase in p_2, market 2 is not heavily influenced by the price change in market 1. Traders in market 1, however are still very

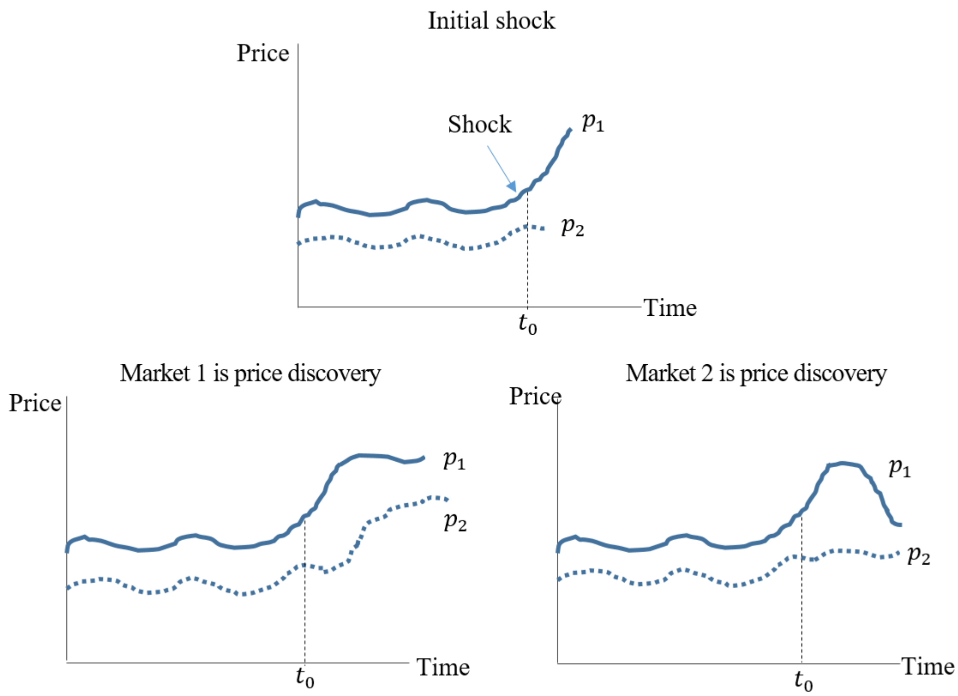

Figure 4. Price discovery illustration

concerned with p_2. Although the shock in market 1 causes a temporary increase in p_1, the price in market 2 indicates to traders in market 1 that p_1 is too high. This information is incorporated into market negotiations, and, over time, p_1 returns to its proper level, slightly higher than p_2 . This change is shown in the third panel in Figure 4.

Vector Error Correction Model and Price Discovery

Vector Error Correction Model

Since price discovery is about an identical/similar asset or commodity traded in different markets, a cointegration framework is adopted. The conventional price discovery measure used in the literature simply compared the speed of adjustment coefficients in the (bivariate) Vector Error

Correction Model (VECM) as a share of the total adjustment[2], which is developed in Gonzalo and Granger (2005) and expanded in Theissen (2002). This measure is used by Schwarz and Szakmary (1994), Foster (1996), Theissen (2002), Eun and Sabherwal (2003), Thurlin (2009), and Arnade and Hoffman (2015) among others.

As explained Arnade and Hoffman (2015) estimates of (absolute) adjustment rates are related to market efficiency. “Short-run efficiency” is defined by the speed with which a displaced price in market 1 and/or a displaced price in other market returns to the long-run equilibrium relationship between prices 1 and 2 (McKenzie and Holt, 2002) as illustrated in Figure 4. The long-run equilibrium between two prices can be written as follows:

(1) $$p_{1,t} = \beta_2 p_{2,t} + c + u_t \;\Leftrightarrow\; u_t = p_{1,t} - \beta_2 p_{2,t} - c$$

where $p_{1,t}$ and $p_{2,t}$ represent prices in market 1 and 2 at time t, respectively. The term c (constant term) account for differences in these two markets. The term u_t is the (long-run) error, which is zero in equilibrium. If $c = 0$ and $\beta_2 = 1$, then markets are “long-run efficient” (Tomek and Gray, 1970).

The bivariate VECM contains this long-run equilibrium in equation (1) as follows:

(2) $$\begin{aligned}\Delta p_{1,t} &= \alpha_1\left(p_{1,t} - \beta_2 p_{2,t} - c\right) + \sum_{k=1}^{K-1} \gamma_{11,k}\Delta p_{1,t-k} + \sum_{k=1}^{K-1} \gamma_{12,k}\Delta p_{2,t-k} + e_{1,t}\\ \Delta p_{2,t} &= \alpha_2\left(p_{1,t} - \beta_2 p_{2,t} - c\right) + \sum_{k=1}^{K-1} \gamma_{21,k}\Delta p_{1,t-k} + \sum_{k=1}^{K-1} \gamma_{22,k}\Delta p_{2,t-k} + e_{2,t}\end{aligned}$$

[2] Other measures of price discovery used in the literature are the Information Share (IS) of Hasbrouck (1995) and the Component Share (CS) of Booth et al. (1999), Chu et al. (1999), and Harris et al. (2002). The IS measures each market's relative contribution to the variance of the efficient price, while the CS decomposes the common efficient price into a weighted average of observed market prices, and measures each market's contribution to the common efficient price. Both IS and CS are based on the reduced-form “forecasting errors” in a Vector Error Correction Model (VECM) (Kim, 2011).

where $\Delta p_{i,t}$ represent the differenced prices. Coefficients α_1 and α_2 are the adjustment rate parameters and they represent the speed with which a displaced p_1 (or p_2) returns to its long-run equilibrium. Using vectors and matrices, equation (2) is simplified as follows

$$(3) \qquad \Delta\mathbf{p}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{p}_{t-1} - \mathbf{c}) + \sum_{k=1}^{K-1} \boldsymbol{\Gamma}_k \Delta\mathbf{p}_{t-k} + \mathbf{e}_t = \boldsymbol{\Pi}(\mathbf{p}_{t-1} + \boldsymbol{\mu}) + \sum_{k=1}^{K-1} \boldsymbol{\Gamma}_k \Delta\mathbf{p}_{t-k} + \mathbf{e}_t$$

where $\boldsymbol{\alpha}' = [\alpha_1, \alpha_2]$ is 2×1 vector of adjustment coefficients (speed of adjustment), $\boldsymbol{\beta}' = [1, \beta_2]$ is 2×1 cointegrating vector, $\mathbf{c}$ is an intercept, $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, and $\boldsymbol{\mu} = -\boldsymbol{\Pi}\mathbf{c}$. K is the lag-length of the vector autoregression (VAR) and $\mathbf{e}_t$ is the reduced-form shock.

Following Gonzalo and Granger (1995), the *relative ratio* of the adjustment coefficients is defined by

$$(4) \qquad \theta_1 = \frac{|\alpha_2|}{|\alpha_1| + |\alpha_2|}, \quad \theta_2 = \frac{|\alpha_1|}{|\alpha_1| + |\alpha_2|}, \text{ and } \theta_1 + \theta_2 = 1.$$

A high (low) θ_i indicates a low (high) α_i, which in turn implies that market i slowly (quickly) responds to an unpredicted shock in the system; therefore market i is (not) the price discovery reference market. If $\theta_1 = \theta_2 = 0.5$, both markets contribute equally to the price discovery process; i.e. both markets move at similar rate toward the long-run equilibrium.

Stationary Test

Stationarity is an essential underlying concept for time-series econometric analysis. Most of price data are not stationary enabling to estimate VECM. Let p_t be the cattle price at time t. The cattle price p_t is stationary if its first two moments are finite and constant over time, meaning that neither the mean nor the autocovariance depend on the date t (Hamilton 1994). Mathematically, $\mathrm{E}(p_t) = \mu$ and $\mathrm{E}\big[(p_t - \mu)(p_{t+j} - \mu)\big] = \gamma_j$ for all t and j, where γ_j is the autocovariance between p_t and p_{t+j}.

Stationarity of the time-series data is important because parameter estimates are not consistent when series are non-stationary. A property of a stationary process is that it is mean-reverting, indicating it will fluctuate around its mean, and parameter estimates remain consistent. In addition, a non-stationary series causes a spurious regression problem. If two variables are trending over time, a regression of one on the other could have a high R^2 value even if the two series are unrelated. To test whether the time-series data is stationary, the unit root test is used. Equation (5) is a possible random walk model:

$$p_t = \rho p_{t-1} + u_t \tag{5}$$

where u_t is a white noise error term. When $\rho = 1$, it becomes a random walk model without drift, implying the existence of a unit root and indicating a non-stationary stochastic process. If we regress p_t on its lagged value p_{t-1} and the estimated ρ is not statistically different from 1, then p_t is non-stationary.

The Dickey-Fuller (ADF) and the Phillips-Perron (PP) (Phillips and Perron, 1988)tests are well-known and commonly applied methods to test for the unit root. In these tests, the null hypothesis is that the series has a unit root, i.e., H_0: $\rho = 1$. If the null hypothesis is rejected, then the data series is stationary. The DF test is based on estimating the equation (5) by ordinary least, and thus it may suffer from a serial correlation problem. The Augmented DF test adds the lagged difference terms of the dependent variable to adjust for the serial correlation. The PP test uses an alternative method to account for the serial correlation problem. The PP test uses Newey–West (1987) standard errors to deal with serial correlation problem in the innovation term. The lag length (or bandwidth) for the PP test is determined using the formula $4\left(\frac{T}{100}\right)^{\frac{2}{9}}$ where T is the sample size as suggested in Newey and West (1994). However, choosing the lag length this way is not necessarily optimal (Hoechle 2007). In this study the PP test is implemented for the unit root test through STATA software.

Deciding Optimal Lag Length

To build VECM model in equation (2), K, the optimal lag length of the (initial) VAR model, should be decided. When too many lags are applied, the error in the forecasts will be bigger. On the contrary, if too few lags are applied, this could leave out relevant information. Therefore, applying optimal lag length is important. The optimal lag length can be determined by minimizing the following information criteria: Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), and the Hannan-Quinn Criterion (HQIC). These information criteria contain both the goodness of fit of the model and the complexity of the model. Each of the information criterions are defined as follows:

$$\text{AIC} = -2\left(\frac{LL}{T}\right) + \frac{2t_K}{T}$$

$$\text{SIC} = -2\left(\frac{LL}{T}\right) + \frac{\ln T}{T} t_K \qquad (6)$$

$$\text{HQIC} = -2\left(\frac{LL}{T}\right) + \frac{2\ln(\ln T)}{T} t_K$$

where t_K is the total number of parameters in the model (number of coefficients to be estimated) and LL is the value of log likelihood function, and T is the number of observations. The first term, $-2\left(\frac{LL}{T}\right)$ indicates an estimate of the deviance of the fit of the model, and the second term of each equation indicates the degree to which the number of model parameters is being penalized. In this study, SIC is used.

Specifying the Cointegrating Rank

If the price series under consideration are I(1), a VECM in equation (2) is the suitable modeling framework. The number of cointegration relationship, in bivariate case we have at most one cointegration vector, can be investigated based on likelihood ratio (LR)-type tests (Lütkepohl and Krätzig, 2004). The hypothesis to test is

$$H_0(0)\colon \text{rank}(\mathbf{\Pi}) = 0 \quad vs.\ H_1(0)\colon \text{rank}(\mathbf{\Pi}) > 0, \qquad (7)$$
$$H_0(1)\colon \text{rank}(\mathbf{\Pi}) = 1 \quad vs.\ H_1(0)\colon \text{rank}(\mathbf{\Pi}) > 1$$

The testing sequence terminates, and the corresponding cointegrating rank is selected when the null hypothesis cannot be rejected for the first time (Lütkepohl and Krätzig, 2004). Johansen (1995) provides critical values for the LR test, which is known as the trace test. The test statistics is of the form

$$\text{LR}(r_0) = -T \sum_{j=r_0+1}^{K} \log\left(1 - \lambda_j\right), \qquad (8)$$

where $\text{LR}(r_0)$ is the LR test statistics for testing whether $\text{rank}(\mathbf{\Pi}) = r_0$ versus $\text{rank}(\mathbf{\Pi}) > r_0$ and the λ_j are the eigenvalues of $\mathbf{\Pi}$.

Problem of Bivariate VECM and Price Discovery

The Achilles heel to price discovery using the speed of adjustment coefficients in with the VECM outlined in equations (2) and (4) is its inability to compare more than two variables at a time. This is due to the fact that there must be one cointegrating relationship between the relevant variables. We may arrange price discovery according to relative magnitude of α_i for multivariate case with one cointegrating relation but it is not clear on how to calculate θ_i (Kim, 2011, pp. 50-51). In addition, there might be more than one cointegrating relationship in the multivariate case. For n time series there might be as many as $n-1$ cointegrating relationships. Since there is no way of choosing in this latter case which of these possible cointegrating relationships to use in the model (Kim, 2011), one cannot use this approach to compare more than two series at a time. This is a limitation when examining markets of a commodity, which almost always involves more than two markets such as cattle markets.

Cluster Analysis and Tournament

Justification of Tournament Approach

As noted when there are multiple markets, this technique, equation (4), may not be applicable. The tournament approach using the cluster analysis provides a way to overcome this problem as demonstrated in Kim, Tejeda, and Wright (2016). The cluster analysis divides cattle price data into groups (clusters) that are meaningful, useful or both (Tan, Steinbach, and Kumar, 2006). It implements a hierarchy using various techniques, for example, pairwise distance between all data points. Once the cluster analysis provides the resulting cluster(s), the *sequential* VECM approach (similar to a tournament) is applied and we can identify the price discovery region(s) for each hierarchy. The main contribution of this paper is the development of a technique that allows the researcher to compare many markets at once.

Early attempts at overcoming this problem included a round robin tournament approach. This means using the error correction model to compare each series to every other series. This method is inefficient, and labor intensive since a study with many variables, involved many comparisons. An example of this approach is Oellermann, Brorsen, and Farris (1989). This paper considered four markets, labeled feeder futures, feeder cash, live cattle cash, and live cattle futures. The authors compared four pairs of variables: (1) feeder cash vs. feeder futures, (2) feeder cash vs live cattle futures, (3) feeder cash vs live cattle cash, and (4) feeder futures vs live cattle futures (Oellermann, Brorsen, and Farris, 1989). Note this is technically not quite a round robin tournament since there are some possible combinations that were not included. These include (5) labeled feeder futures vs live cattle futures and (6) live cattle futures vs live cattle cash.

Another, more efficient, approach is to use a single elimination tournament. Using this approach in a study with n markets allows us to reduce the number of comparisons to $n - 1$. For example, if one were to do a study including prices from 30 different markets, one would only have to run the 29 bivariate VECMs. A downside to a single elimination tournament is we cannot say who the second strongest competitor is. Just because a competitor losses in the final round does not mean it is second strongest. The second, third and fourth strongest competitors may have already been matched against the strongest competitor and eliminated by the time the final round is played.

Sports tournaments often address this issue through the use of double and triple elimination tournaments. In a double elimination tournament, one must lose twice before being eliminated. The tournament is designed so that if two teams play each other once, they don't play each other again until the final round. This ensures the second strongest team makes it to the final round before being eliminated. The triple elimination tournament works on the same principle except it ensures the top three teams make it to the final stages of the tournament. Unfortunately, a double elimination tournament requires twice as many matches as a single elimination tournament. A triple elimination tournament requires three times as many matches, etc. Rather than use a double or triple elimination tournament, along with the computational load that would entail, we will use a single elimination tournament and we will attempt to offset the deficiencies of this approach by incorporating a cluster analysis. We will use this data mining tool to segregate our variables into groups or clusters, allowing us to identify 'winners' in sub-categories of the U.S. cattle market.

Cluster Analysis[3]

To perform a cluster analysis means nothing more than dividing data into groups. This is usually done to make the data more meaningful, useful, or both. (Tan, Steinback, and Kumar, 2006, p. 487). During this process, of grouping a set of data or other, objects are sorted into multiple groups so that objects within a cluster have high similarity, but are less similar to objects in other

[3] This section is developed heavily based on Chapter 10 in Han, Kamber, and Pei (2011) and Chapter 8 in Tan, Steinback, and Kumar (2006).

clusters. Sameness can be based on whatever factors the researcher considers relevant. Quite often, differences involve distance measures. Regardless of difference criteria, the goal of this process is to create clusters where the objects within a group are similar to one another and different from the objects in the other groups. Greater similarity within a group and larger differences between groups is usually preferred, since this leads to more distinct clustering.

With time series data, like that used in this study, the most popular distance measure is Euclidean distance.[4] Let $p_{i,t}$ be the (cattle) price in market $i = 1, \cdots, n$. The Euclidean distance between markets i and j is defined as follows:

$$d(i,j) = \sqrt{\sum_{t=1}^{T} \left(p_{i,t} - p_{j,t}\right)^2} \quad (9)$$

Note that the Euclidean distance is nonnegative, the distance of a market to itself is 0, i.e., $d(i,i) = 0$ and the distance is a symmetric, i.e., $d(i,j) = d(j,i)$.

Clustering comes in different flavors, including nested and unnested, i.e., partitional or hierarchical. In a partitional clustering, the set of data objects is sorted into non-overlapping subsets (clusters) so that each object is in one, and only one subset. To obtain a hierarchical clustering, we permit clusters to have subclusters, which are a set of nested clusters. Graphical representations of clusters resemble a tree.

A hierarchical clustering method works by grouping markets into a tree of subclusters. We can further divide hierarchical clustering methods into two types i) agglomerative hierarchical clustering and ii) divisive hierarchical clustering. The first type, Agglomerative hierarchical

[4] Another well-known metrics are Manhattan distance, $d(i,j) = \sum |p_{i,t} - p_{j,t}|$ and Minokowski distance, $d(i,j) = \left(\sum \left|p_{i,t} - p_{j,t}\right|^m\right)^{1/m}$, where m is a positive integer. Minokowski distance represent the Manhattan distance when $m = 1$ and Euclidean distance when $m = 2$.

clustering, is based on a bottom-up strategy. It usually starts by letting one object form a cluster and forms successively larger clusters around the original one, until all objects have been sorted. The single cluster is the starting point for the hierarchy. A divisive hierarchical clustering method works in the opposite direction. It starts by placing all the objects in one group. The group is then split into several smaller groups. These smaller groups are then each partitioned into even smaller groups. This process is repeated until objects within each group are similar enough to each other.

Figure 5 shows of agglomerative hierarchical clustering method and a divisive hierarchical clustering method, on a data set of five objects, $\{a, b, c, d, e\}$. Initially, the agglomerative method, places each object into a cluster of its own. The clusters are then merged step-by-step according to some criterion. For example, clusters C_1 and C_2 may be merged if an object in C_1 and an object in C_2 form the minimum Euclidean distance between any two objects from different clusters. This is a single-linkage approach in that each cluster is represented by all the objects in the cluster, and the similarity between two clusters is measured by the similarity of the closest pair of data points belonging to different clusters.

The cluster-merging process repeats until all the objects are eventually merged to form one cluster. The divisive method, proceeds in the contrasting way. All the objects are used to form one initial cluster. The cluster is split according to some principle such as the maximum Euclidean distance between the closest neighboring objects in the cluster. The cluster-splitting process repeats until, eventually, each new cluster contains only a single object.

Tree structure called a dendrogram is commonly used to represent the process of hierarchical clustering. It shows how objects are grouped together (in an agglomerative method) or partitioned (in a divisive method) step-by-step. Figure 6 shows a dendrogram for the five objects presented in Figure 5, where $l = 0$ shows the five objects as singleton clusters at level 0. At $l = 1$, objects a and b are grouped together to form the first cluster, and they stay together at all subsequent levels. We can also use a vertical axis to show the similarity scale between clusters. For example, when the similarity of two groups of objects, $\{a, b\}$ and $\{c, d, e\}$ is roughly 0.16 (similarity scale on the right in Figure 6), they are merged together to form a single cluster.

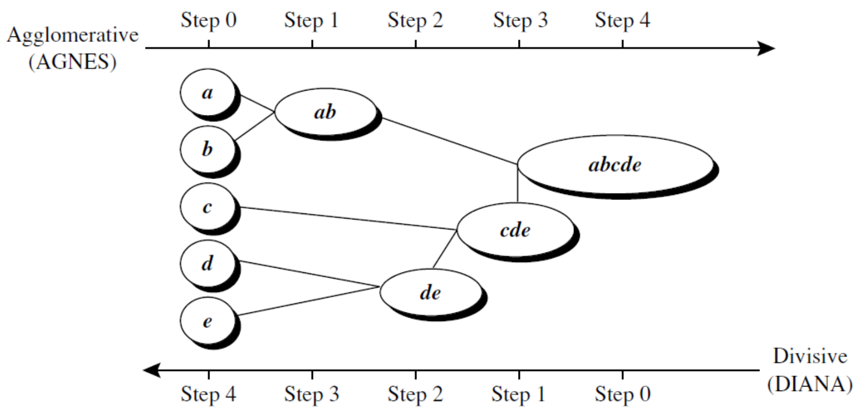

Figure 5. Agglomerative and divisive hierarchical clustering
Reproduced from Figure 10.6 in Han, Kamber and Pei (2011), p. 460

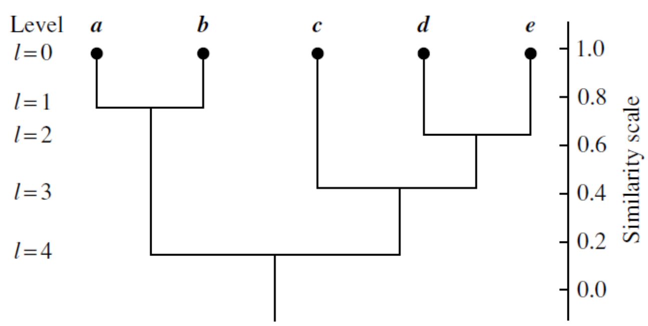

Figure 6. Dendrogram representation for hierarchical clustering
Reproduced from Figure 10.6 in Han, Kamber and Pei (2011), p. 460

Whether using an agglomerative method or a divisive method, a core need is to measure the distance between two clusters, where each cluster is generally a set of objects. Four widely used measures for distance between clusters are as follows, where $|p - p'|$ is the distance between two markets, m_i is the mean for cluster C_i and n_i is the number of observations in C_i. They are also known as linkage measures

$$\text{minimum distance:} \quad d_{min}(C_i, C_j) = \min_{p \in C_i, p' \in C_j} \{|p - p'|\}$$

$$\text{maximum distance:} \quad d_{max}(C_i, C_j) = \max_{p \in C_i, p' \in C_j} \{|p - p'|\}$$

$$\text{mean distance:} \quad d_{mean}(C_i, C_j) = |m_i - m_j| \tag{10}$$

$$\text{average distance:} \quad d_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i, p' \in C_j} |p - p'|$$

In this research, `hclust` function in R software is utilized that uses the maximum distance, $d_{max}(C_i, C_j)$, to measure the distance between clusters. It is sometimes called a complete-linkage algorithm. The maximum distance (and minimum distance, too) tend to be sensitive to outliers or noisy data. The use of mean or average distance is a compromise this problem.

DATA

This study uses weekly cattle price data from January of 2001 through October of 2016. All of cattle cash and futures price data are compiled from Livestock Marketing Information Center (LMIC). The live fed cattle data was collected from all auctions within a region. Regions studied include Texas (TX), Oklahoma (OK), Kansas (KS), Nebraska (NE), Colorado (CO), and Iowa (IA). An average of these auctions was taken to produce a single series for steers and heifers, representing the fed cattle price for that region. All weights and grades of cattle were included in this calculation. For live feeder cattle, only calves from 500 to 900 pounds were included. Data from all the auctions in the relevant geographical area were included. Regions studied include CO, KS, Missouri (MO), Montana (MT), NE, OK, South Dakota (SD), and TX. The raw data gave an average price for steers and heifers, divided into categories by hundred pound increments. Head counts from these auctions were not available at the time of this study. So, a simple, unweighted average was calculated for both steers and heifers in the specified weight range.

Unfortunately, both the live fed cattle, and the live feeder cattle series contained missing observations. The data set was made complete by taking an average of the observations above and below the missing point and using this estimate of the missing observation. If there was more than one missing observation in a row, the missing points were estimated to be contained in a line between the closest observable points (linear interpolation). In addition, an extensive number of typos were present in the raw live feeder cattle data. Prices greater than 300 were removed from the data and these missing observations were estimated using linear interpolation. The price level of 300 was chosen to maximize the amount of data retained, while still eliminating outliers and possible mistakes in the raw data.

Boxed beef cutout price is also included which is provided by the LMIC in the form of a weekly weighted average. The raw box beef cutout data had no missing observations, and was used as provided. In addition to these prices, the cattle feeder cattle index is included, too. The feeder cattle index is a seven-day weighted average of the total dollars sold divided by the total pounds during the same period. The USDA-AMS issues daily reports which contain cattle eligible for the index. In total, we have 30 cattle price series which are the combinations of regions, feeder/fed, steer/heifer, and futures prices. Table 1 in Appendix A presents basic statistics and Figures 7 and 8 contain plots of price series.

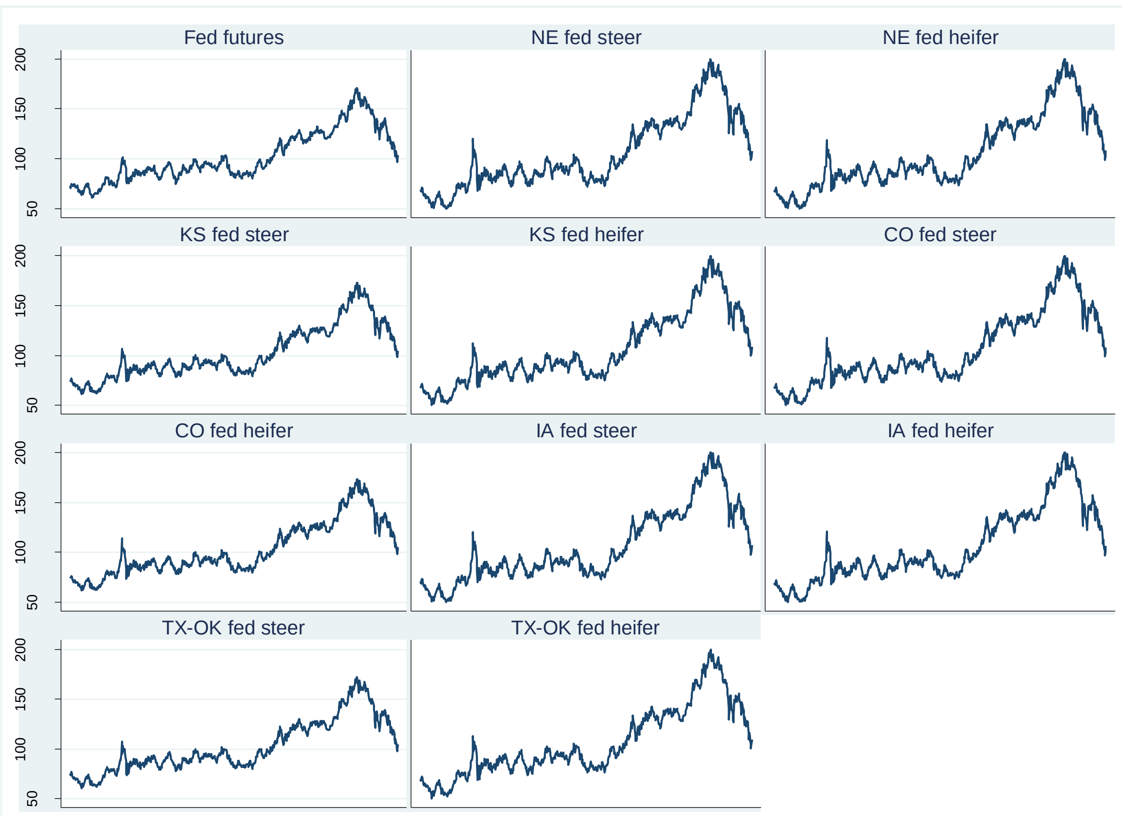

Figure 7. Fed cattle price across regions and types
Source: Livestock Marketing Information Center (LMIC)
State_fed_S = fed cattle steer in state, Statee_fed_H = fed cattle heifer

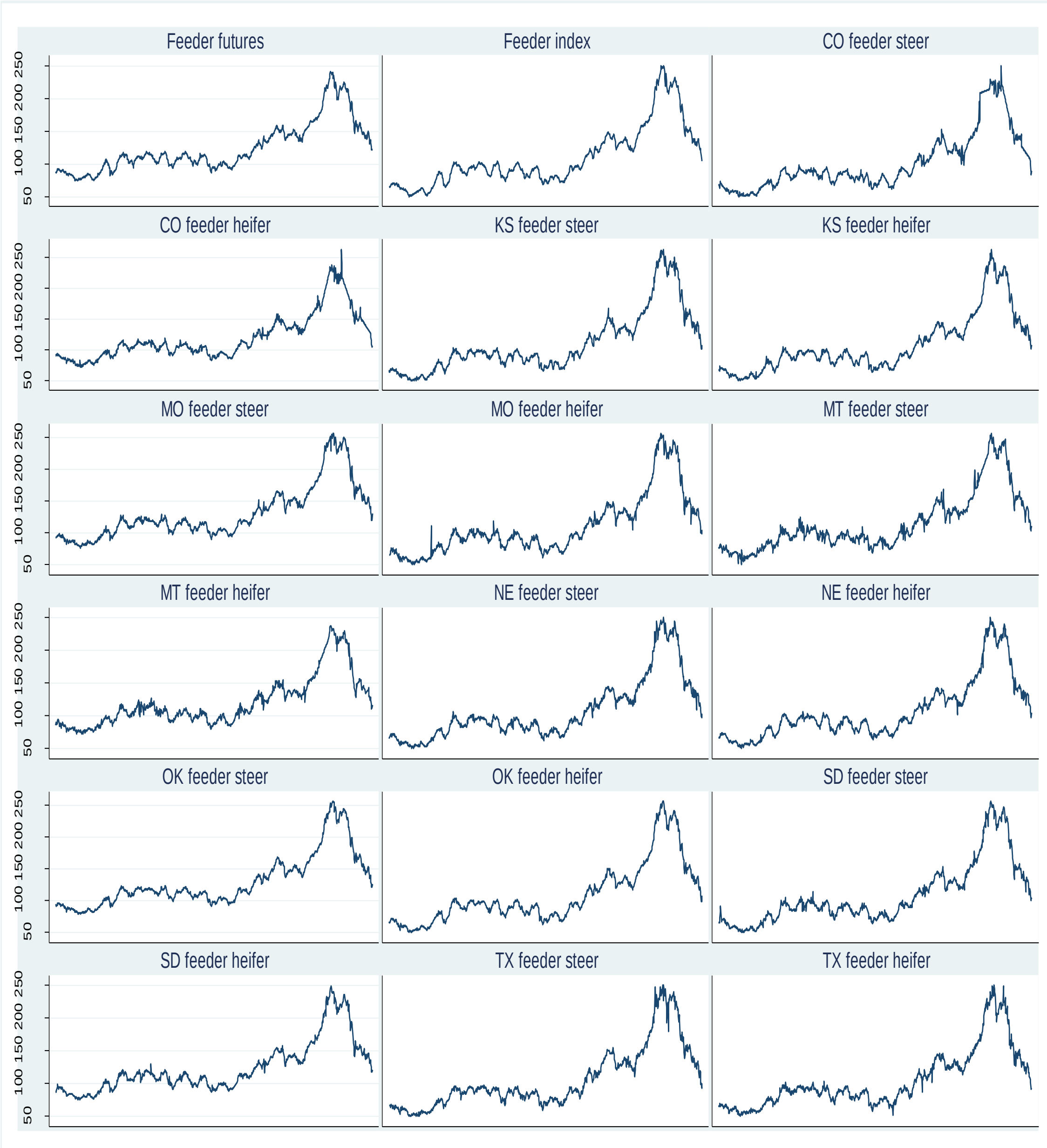

Figure 8. Feeder cattle price across regions and types

Source: Livestock Marketing Information Center (LMIC)

State_feeder_S = feeder cattle steer in state, State_feeder_H = feeder cattle heifer

RESULTS AND DISCUSSION

Cluster Analysis

The cluster analysis provides the following cluster dendrogram in Figure 9 using `hclust` function in R software. As can be seen from the dendrogram, this clustering creates a special kind of tournament bracket. In this bracket, those markets most closely aligned with each other are compared first. The advantage of this approach is that it allows us to draw more conclusions about segments of the market than would be possible if we randomly assigned variables to different places in a single elimination tournament. For instance, all the fed cattle markets are grouped into the same

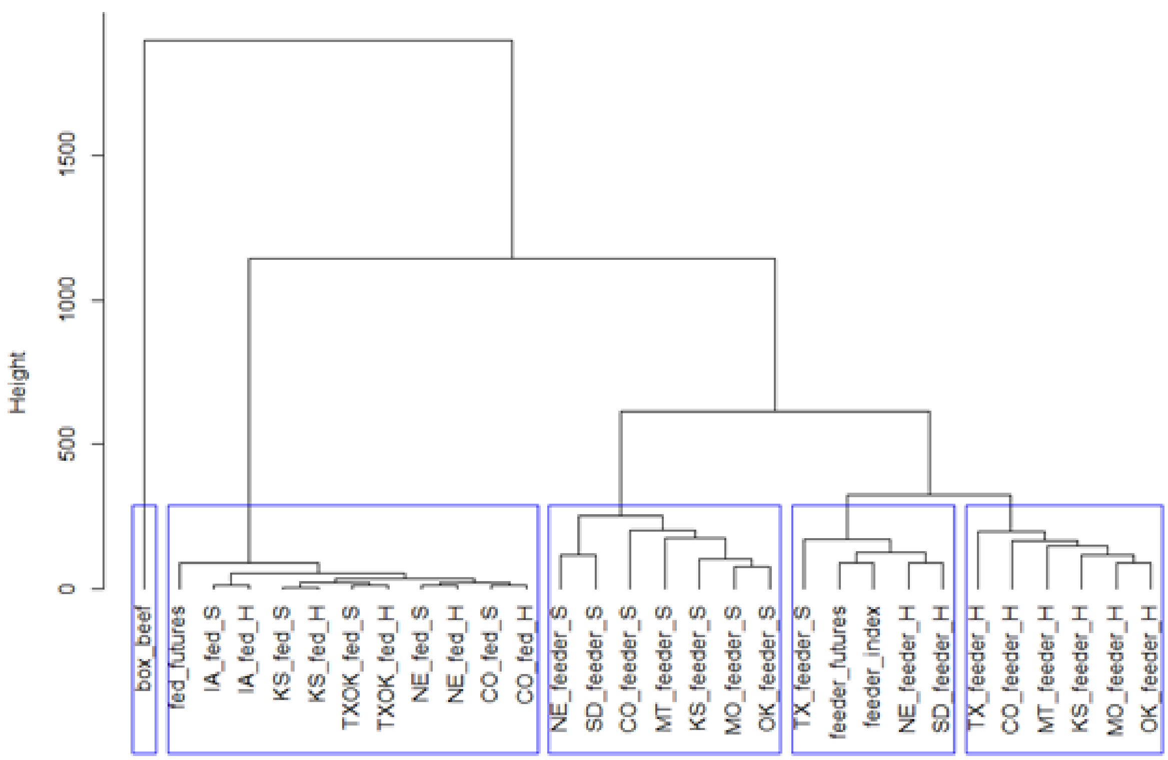

Figure 9. Cluster analysis dendrogram

State_fed_S = fed cattle steer in state, State_fed_H = fed cattle heifer in state, State Name_feeder_S = feeder cattle steer in state, State Name_feeder_H = feeder cattle heifer in state hierarchy (second blue box in Figure 9). This fact allows us to determine the price discovery for the U.S. fed cattle market. The implications of this approach will be made clearer subsequently.

The cluster analysis allows us to draw conclusions about market structure from the height of the clusters that is shown to the left of the dendrogram in Figure 9. As previously explained in equation (9), this number represents the Euclidian distance between markets i and j. We can use this equation (9) to calculate the average difference in price which is,

$$\text{Avg price distance} \approx \sqrt{\frac{d(i,j)^2}{T}} \quad (11)$$

where T is the total number of observations[5].

We can now plug in the appropriate numbers to see how much price variation occurs in the different cattle markets. For instance, we can see from the dendrogram in Figure 9, that the height on the box containing the fed cattle markets (second box in Figure 9) is approximately 100. We have 809 observations. So, the average difference in price for the time periods examined, for the two series furthest apart in the fed market is \$3.52.[6] In contrast, the height for the feeder market is about 700. This gives an average difference, within the feeder market of \$24.60. This substantial difference demonstrates the consolidation that has occurred in the fed cattle market. However, there are still many producers of feeder cattle. This feeder production also occurs in many areas of the U.S. Consequently, there is much more price variation within this market. The fed cattle market,

---

[5] We can find $\sum_{t=1}^{T}\left(p_{i,t} - p_{j,t}\right)^2 = d(i,j)^2$ from equation (9) and the average of *squared* differences in price is given by $\frac{d(i,j)^2}{T}$. The average difference in price is given in equation (11) by taking a square root. Note that it is an approximation of the average differences in price not the actual average ignoring all the cross products of $p_{i,t}$ and $p_{j,t}$.

[6] $\sqrt{\frac{d(i,j)^2}{T}} = \sqrt{\frac{100^2}{809}} \approx 3.52$

on the other hand, is dominated by a relatively few number of very large producers. This has resulted in much less price variation within this market.

Although the original intent of this paper was to determine the price discovery for the entire U.S. cattle market, the cluster analysis demonstrates the submarkets within this industry are separate enough from each other that they should be viewed as distinct markets, rather than subcategories within the same market. The distance, for example between the fed and feeder markets is approximately 1200. This translates to an average difference at each time period of $42.16. The distance between the boxed beef market, and the other markets studied is about 2000. This translates into an average price spread of $70.27. Although there are obvious connections between these markets, these large price spreads demonstrate the significant separations between these markets. Due to these results, we determine the price discovery for the fed market, as well as the feeder market. We do not, however, attempt to determine a winner between the two. From this point forward, we also view boxed beef as a separate market, excluding it from our analysis.

The submarkets identified by the cluster analysis include the following:

(1) *Fed cattle market*: this market includes all live fed cattle series, as well as fed cattle futures (second blue box in Figure 9).

(2) *Feeder cattle market*

   a. *Feeder steer market*: this market includes all feeder steers except those from Texas (third blue box in Figure 9).

   b. *Feeder heifer and feeder futures market*: this market includes all feeder heifers, feeder cattle futures, the feeder cattle index, and feeder steers from Texas (fourth and fifth blue box in Figure 9).

We will examine these submarkets and then compare the winners from each category.

A VECM for each pair of these submarkets 9 is estimated, we then determine the price discovery region using values of θ_i from equation (4), and the leader of each binary contest goes up one hierarchy. Another VECM is estimated for the new (hierarchical) pair of markets and then the price discovery region is identified until we have the final “winner”.

Stationarity Tests and Cointegration Tests

Before moving to Johansen’s test to check if there exists one cointegrating relationship among pairs of series, stationary tests for each price series were conducted. This study utilizes the Phillip Perron test to check for stationarity of each variable. Table 3 in Appendix C gives the results. As expected all the cattle prices are non-stationary. All matched series were tested for cointegration using the Johansen’s trace test (all were found to be cointegrated) and the results of tests are given in Table 4 in the appendix D. The fact that all these pairs are cointegrated makes them suitable for analysis with the error correction model.

Fed Cattle Market

The first group we will examine is the fed cattle market. Figure 10 summarizes the results. Figure 10 includes the name of the series being compared, followed by θ_i in equation (4) using the adjustment coefficients from the corresponding VECM. For instance, the bottom left corner of the diagram shows the comparison between CO fed steer and CO fed heifer. The θ value for the CO fed steer is .299 and the θ value for CO fed heifer is .701. Clearly, this indicates CO fed heifer is the winner for this round in the tournament. Consequently, it moves forward to compete against NE fed steer which is a winner of NE fed steer and NE fed heifer match.

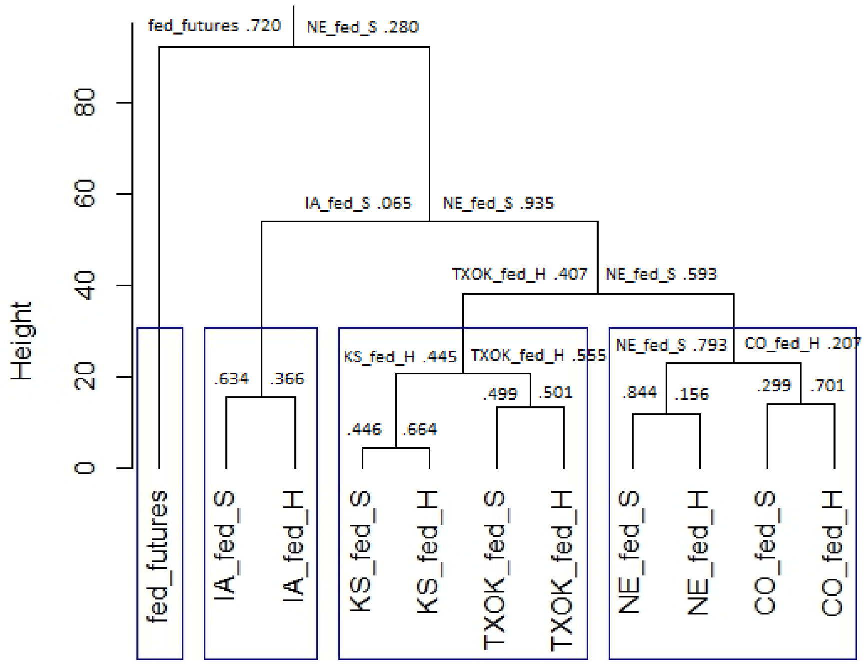

Figure 10. Fed cattle tournament result

State_fed_S = fed cattle steer in state, State_fed_H = fed cattle heifer in state

Although this is a single elimination tournament, comparison of pairs of series through the structure provided by the cluster analysis allows us to conclude NE fed steer is the price discovery when only live fed cattle markets are considered, i.e., as shown in Figure 10, NE fed steer play an important part in the price discovery process. However, when we compare fed futures to NE fed steer, we see the price is actually discovered in the fed futures market. This confirms results from an extensive literature on futures markets going back at least to the early 1940s (Working, 1942).

Feeder Cattle Heifers

Feeder Heifer and Feeder Futures

The second group includes the feeder heifers, feeder futures, the feeder index and TX feeder steers (Figure 11). This bracket exemplifies some of the weaknesses of our approach. The feeder index contains more information than any of the particular live feeder markets. Yet, because it lies closest to feeder futures, it is eliminated without being compared to any of the live feeder markets. One might be tempted to conclude from the diagram that KS feeder heifers play a prominent role in the price discovery process. This, however, would be a mistake. A closer examination of the bracket reveals KS feeder steers could actually be less important in the price discovery process than the feeder index, feeder futures, NE feeder heifers, SD feeder heifers, and TX feeder steers. This would place KS feeders somewhere in the middle of the pack, rather than in second place. The one conclusion we can draw is that feeder futures is the price discovery in this subdivision of the U.S. cattle market.

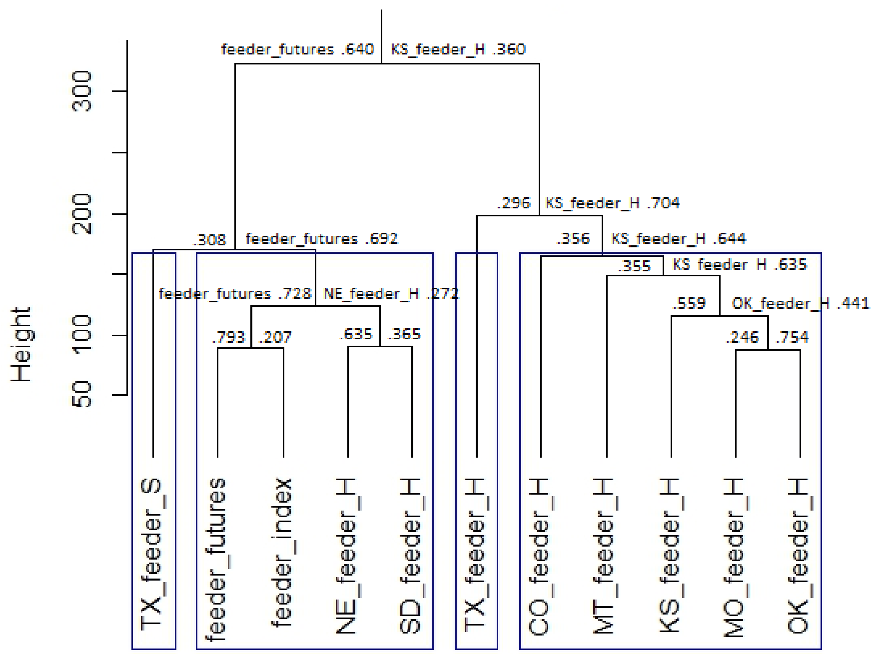

Figure 11. Feeder cattle heifers tournament result

State Name_feeder_S = feeder cattle steer in state, State Name_feeder_H = feeder cattle heifer in state

Feeder Cattle Steers

The third group includes all the feeder steers except those from TX. As the diagram in Figure 12 shows, CO feeder steers are the price discovery out of the variables compared.

It is still necessary to compare the winner from the feeder cattle heifers tournament with the winner from the feeder cattle steers tournament. When we do this, we find the theta value for Colorado feeder steers to be .380 and the theta value for feeder futures to be .620. As explained in the section on fed cattle markets, this result confirms a large body of research showing the correct price for commodities is discovered in the futures market.

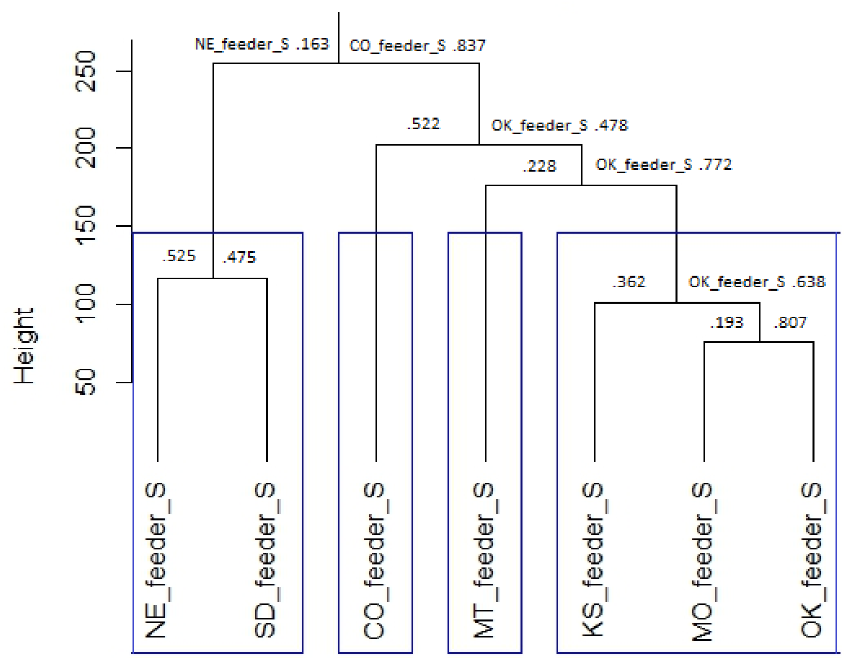

Figure 12. Feeder cattle steers tournament result

State Name_feeder_S = feeder cattle steer in state, State Name_feeder_H = feeder cattle heifer in state

SUMMARY AND CONCLUDING REMARKS

Price discovery in cattle markets is a topic of interest for professionals working in the industry as well as for economists. This research examines price discovery in the U.S. cattle market across regions and types. It improves on previous attempts by incorporating a tournament approach with a cluster analysis, which allows more markets to be considered in the price discovery than was previously considered feasible.

A preliminary chapter was included, giving a brief overview of the U.S. cattle market. It illustrates why specialization has occurred with the market and shows the roles each sub-market plays in the beef production process. This chapter also explained how the variables used in this study, such as the feeder index, box beef cutouts, are calculated.

The next chapter presents the price discovery concept. It gives a history of the literature on the subject, including the econometric methods that have been developed to empirically investigate price discovery. The error correction model was presented. The necessary requirements to apply this model were also discussed. As was pointed out, this model is a useful tool for quantifying the price discovery process. It has the major shortcoming, however, of only being able to compare two series at a time. This is due to the fact that series compared with the model must have only one cointegrating relationship. The number of possible cointegrating relationships is one less than the number of variables considered. Thus, with more than two variables, there may be multiple possible cointegrating relationships. Since it is unclear which of these relationships should be used, we are unable to use the model with three or more markets.

The major contribution of this paper is the tournament approach with a cluster analysis. The paper discusses different kinds of tournaments that can be used, along with their strengths and weaknesses. In the end, a cluster analysis approach was adopted. An extensive discussion on different approaches to cluster analysis was presented. The use of a cluster analysis allows to avoid the computation burden imposed by a round robin tournament and still tease out more information than would be possible with a random single elimination tournament.

The variables were tested and it was found they all met the criteria for the error correction model. The cluster analysis was conducted, and the resulting dendrogram was used to construct a single elimination tournament. The tournament was conducted by using the estimated (bivariate) error correction models to determine winners and losers. The final results of the tournament confirmed theoretic expectations, as well as previous empirical work done on this topic. This study shows the proper price for cattle is discovered in the futures market.

This study addresses past deficiencies in price discovery research; however, there is still much work to be done. Although the single elimination tournament, in conjunction with the cluster analysis, allows more variables to be taken into consideration than previously, it suffers from the limitations inherent in single elimination tournament design. If the second strongest competitor is paired against the strongest early in the tournament, it will be eliminated long before the final round. Thus, relative ranking of all variables is not possible with this approach. Although computationally burdensome in the past, modern computing power now makes a round robin approach feasible. Using the error correction model to make all possible comparisons between the variables and counting the number of wins would allow one to present a complete ranking of the importance of all the variables compared in the price discovery process. This, and other further refinements should provide avenues for further exploration of the price discovery concept.

REFERENCES

Arnade, C. and L. Hoffman. 2015. The Impact of Price Variability on Cash/Futures Market Relationships: Implications for Market Efficiency and Price Discovery. *Journal of Agricultural and Applied Economics* 47: 539-559.

Booth, T., R. So, and Y. Tse. 1999. Price Discovery in the German Equity Index Derivatives. *Journal of Futures Markets* 19(6): 619-643.

Chu, Q., W. Hsieh, and Y. Tse. 1999. Price Discovery on the S&P 500 Index Markets: An Analysis of Spot Index, Index Futures, and SPDRs. *International Review of Financial Analysis* 8(1): 21-34.

Eun, C., and S. Sabherwal. 2003. Cross-border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks. *The Journal of Finance* 58: 549-575

Foster, A.J. 1996. Price Discovery in Oil Markets: A Time Varying Analysis of The 1990-1991 Gulf Conflict. *Energy Economics* 18: 231-246.

Figuerola-Ferretti, I. and J. Gonzalo. 2010. Modelling and Measuring Price Discovery in Commodity Markets. *Journal of Econometrics* 158: 95–107.

Gonzalo, J. and C. Granger. 1995. Estimation of Common Long-memory Components in Cointegrated Systems. *Journal of Business & Economic Statistics* 13: 27–35.

Han, Jiawei, Jian Pei, and Micheline Kamber. (2011) *Data mining: concepts and techniques*. Elsevier.

Harris F., T.H. McInish, and R.A. Wood. 2002. Security Price Adjustments across Exchanges: An Investigation of Common Factor Components of Dow Stocks. *Journal of Financial Markets* 5(3): 277-308.

Hasbrouck, J. 1995. One Security, Many Markets: Determining the Contribution to Price Discovery. *Journal of Finance* 50: 277-308.

Hoechle, Daniel. 2007 Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence. *Stata Journal* 7.3: 281.

Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* New York: Oxford University Press.

Joseph, K., P. Garcia, and P.E. Peterson. 2013. Price Discovery in the U.S. Fed Cattle Market. Working paper, available at http://www.farmdoc.illinois.edu/nccc134/conf 2013/pdf/Joseph Garcia Peterson NCCC- 134 2013.pdf

Kim, C-H. 2011 *Essays on Price Discovery*, Dissertation, Department of Economics, University of Washington.

Kim, M-K., H.A. Tejeda, and J. Wright. 2016. Price Discovery in the U.S. and World Rice Markets using a Tournament Approach. Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, MA, July 31-August 2

Koontz, S.R., P. Garcia, and M.A. Hudson. 1990. Dominant-Satellite Relationships between Live Cattle and Cash Futures Markets. *The Journal of Futures Markets* 10(2): 123-136.

Lee, A.C. and M-K. Kim. 2007. Fed Cattle Price Discovery: Directed Acyclic Graph and Time Series Modeling. Selected paper presented at the 2007 American Agricultural Economics Association (AAEA) Annual Meeting, Portland, OR (July).

Lütkepohl, H. and M. Krätzig. 2004. *Applied time series econometrics*. Cambridge university press

McKenzie, A.M. and M.T. Holt. 2002. Market Efficiency in Agricultural Futures Markets. *Applied Economics* 34 (12):1519-1532.

Newey, W.K. and K.D. West. 1987. Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review* 777-787.

Oellermann, C.M., B.W. Brorsen, and P.L. Farris. 1989. Price Discovery for Feeder Cattle. *Journal of Futures Markets* 9: 113–121.

Phillips, P.C.B. and P. Perron. 1988. Testing for a Unit Root in Time Series Regression. *Biometrika* 75(2): 335–346.

Plato, G. and L. Hoffman. 2011. Price Discovery in U.S. and Foreign Commodity Futures Markets: The Brazilian Soybean Example. Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Risk Management, St. Louis, MO, 2011.

Schroeder, T.C. and B.K. Goodwin. 1991. Price Discovery and Cointegration for Live Hogs. *Journal of Futures Markets* 11: 685-696

Schroeder, T.C. 1996. Spatial Fed Cattle Transaction Price Relationships. Hayenga, M.L., S.R. Koontz, and T.C. Schroeder, Ed. *Definition of Regional Cattle Procurement Markets*, GIPSA-RR 96-1: 69-105.

Schwarz, T.V. and A.C. Szakmary. 1994. Price Discovery in Petroleum Markets: Arbitrage, Cointegration, and the Time Interval of Analysis. *Journal of Futures Markets* 14:147–167.

Tan, P-N., M. Steinback, and V. Kumar. 2006. *Introduction to Data Mining*, Pearson Addison Wesley.

Theissen, E. 2002. Price Discovery in Floor and Screen Trading Systems. *Journal of Empirical Finance* 9: 455–474.

Thurlin, A. 2009 *Essays on Market Microstructre: Price Discovery and Informed Trading*. Dissertation, Department of Finance and Statistics, Hanken School of Economics, Helsinki, Finland.

Tomek, W.G. and R.W. Gray. 1970. Temporal Relations among Prices on Commodity Futures Markets. *American Journal of Agricultural Economics* 52(3): 372-380.

United States Department of Agriculture (USDA). 2016. *Overview of the United States Cattle Industry*, National Agricultural Statistics Services, available at http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1648

United States Department of Agriculture (USDA). 2017. *Cattle & Beef*, available at https://www.ers.usda.gov/topics/animal-products/cattle-beef/

Ward, C.E. and T.C. Schroeder. 2002. Price Determination versus Price Discovery. *Managing Today's Cattle Market and Beyond*, available at http://marketing.uwagec.org/MngTCMkt/#update

Womach, J. 2005. *Agriculture: A Glossary of Terms, Programs, and Laws, 2005 Edition*, CRS Report for Congress, Order Code 97-905.

Working, H. 1942. Quotations on Commodity Futures as Price Forecasts. *Econometrica* 10(1): 39-52.

APPENDICES

Appendix A Basic Statistics

Table 1. Basic Statistics (dollars per cwt)

Variables	Mean	Std. Dev.	CV	Min	Max	Autocorr.
feeder_futures	124.42	38.40	30.86	74.30	241.19	0.997
feeder_index	124.66	38.90	31.20	73.93	242.71	0.998
fed_futures	102.11	25.51	24.98	60.64	170.76	0.995
boxed_beef	167.12	38.11	22.80	108.18	263.19	0.993
TXOK_fed_S	102.44	26.15	25.53	60.32	172.00	0.995
TXOK_fed_H	102.51	26.19	25.55	60.10	173.00	0.995
KS_fed_S	102.37	26.24	25.63	61.03	172.94	0.995
KS_fed_H	102.40	26.21	25.59	60.95	172.87	0.995
NE_fed_S	102.36	26.50	25.89	61.84	172.06	0.995
NE_fed_H	102.51	26.51	25.86	61.87	172.44	0.995
CO_fed_S	102.61	26.62	25.94	61.15	173.27	0.994
CO_fed_H	102.68	26.60	25.90	61.66	172.97	0.994
IA_fed_S	101.93	26.30	25.80	61.60	169.93	0.995
IA_fed_H	101.86	26.25	25.77	61.91	170.53	0.995
CO_feeder_S	130.95	42.81	32.70	79.95	274.94	0.994
CO_feeder_H	120.09	37.33	31.08	71.90	263.00	0.994
KS_feeder_S	132.13	42.51	32.17	80.29	261.62	0.996
KS_feeder_H	121.59	38.67	31.80	73.56	248.69	0.997
MO_feeder_S	130.65	42.05	32.19	76.41	255.46	0.996
MO_feeder_H	119.34	37.72	31.61	71.71	234.76	0.993
MT_feeder_S	129.58	42.23	32.59	69.50	260.51	0.992
MT_feeder_H	120.09	38.27	31.87	71.60	237.29	0.993
NE_feeder_S	136.16	43.90	32.24	82.19	268.79	0.995
NE_feeder_H	125.04	39.62	31.69	74.54	249.80	0.995
OK_feeder_S	129.86	41.50	31.96	79.27	256.04	0.998
OK_feeder_H	119.16	37.62	31.57	72.36	234.86	0.997
SD_feeder_S	134.32	43.30	32.24	81.56	266.83	0.995
SD_feeder_H	124.08	39.17	31.57	75.35	248.23	0.995
TX_feeder_S	126.05	41.10	32.60	76.70	250.10	0.994
TX_feeder_H	116.17	37.16	31.99	70.30	236.32	0.993

Source: Livestock Marketing Information Center (LMIC)
State Name_fed_S = fed cattle steer in state, State Name_fed_H = fed cattle heifer, State Name_feeder_S = feeder cattle steer, State Name_feeder_H = feeder cattle heifer

Appendix B Stationarity Test Results

Table 2. Phillips-Perron Stationarity Test Results

Variables	Test statistics Z(rho)[1]	Test statistics Z(t)[1]	5% critical value	MacKinnon approximate p-value for Z(t)[2]
feeder_futures	-3.198	-1.366	-14.10	0.5986
feeder_index	-2.999	-1.325	-14.10	0.6177
fed_futures	-3.994	-1.551	-14.10	0.5083
box_beef	-5.517	-1.732	-14.10	0.4146
TXOK_fed_S	-3.817	-1.493	-14.10	0.5368
TXOK_fed_H	-3.836	-1.496	-14.10	0.5353
KS_fed_S	-3.945	-1.511	-14.10	0.5279
KS_fed_H	-3.943	-1.512	-14.10	0.5277
NE_fed_S	-4.047	-1.524	-14.10	0.5214
NE_fed_H	-4.006	-1.519	-14.10	0.5239
CO_fed_S	-4.038	-1.525	-14.10	0.5211
CO_fed_H	-4.075	-1.531	-14.10	0.5183
IA_fed_S	-3.877	-1.488	-14.10	0.5395
IA_fed_H	-3.919	-1.499	-14.10	0.5341
CO_feeder_S	-3.185	-1.316	-14.10	0.6220
CO_feeder_H	-3.387	-1.349	-14.10	0.6065
KS_feeder_S	-3.076	-1.328	-14.10	0.6163
KS_feeder_H	-3.265	-1.363	-14.10	0.6000
MO_feeder_S	-3.313	-1.378	-14.10	0.5926
MO_feeder_H	-3.775	-1.462	-14.10	0.5521
MT_feeder_S	-3.828	-1.451	-14.10	0.5577
MT_feeder_H	-3.866	-1.474	-14.10	0.5461
NE_feeder_S	-3.159	-1.346	-14.10	0.6080
NE_feeder_H	-3.505	-1.408	-14.10	0.5784
OK_feeder_S	-2.838	-1.277	-14.10	0.6396
OK_feeder_H	-3.034	-1.321	-14.10	0.6197
SD_feeder_S	-3.371	-1.386	-14.10	0.5887
SD_feeder_H	-3.509	-1.416	-14.10	0.5748
TX_feeder_S	-3.355	-1.369	-14.10	0.5972
TX_feeder_H	-3.799	-1.445	-14.10	0.5607

1. Z(rho) is the Phillips-Perron ρ test statistic, and Z(t) is the Phillips-Perron adjusted τ test statistic for the coefficient of ρ in the equation 5. 2. P-values are based on the MacKinnon approximate for Z(t). There was no standard asymptotic distribution to test a unit root. MacKinnon (1994) calculated asymptotic distribution function, and with the result of the study, *P*-value can be applied.
3. If null hypothesis is rejected, then the series is stationary.

Appendix C Johansen's Trace Tests Results – All the Pairs of Prices

Table 3. Johansen's Trace Tests

Price pair	Johansen's Trace Tests for Cointegration					Decision
	Rank	LR	p-value	95%	99%	
MO_feeder_H, OK_feeder_H	0	190.56	0.000	20.16	24.69	R
	1	1.97	0.7793	9.14	12.53	FR
KS_feeder_H, OK_feeder_H	0	178.1	0.000	20.16	24.69	R
	1	1.96	0.7814	9.14	12.53	FR
MT_feeder_H, KS_feeder_H	0	194.67	0.000	20.16	24.69	R
	1	1.65	0.8376	9.14	12.53	FR
CO_feeder_H, KS_feeder_H	0	87.85	0.000	20.16	24.69	R
	1	1.57	0.85	9.14	12.53	FR
TX_feeder_H, KS_feeder_H	0	108.52	0.000	20.16	24.69	R
	1	1.78	0.8143	9.14	12.53	FR
NE_feeder_H, SD_feeder_H	0	197.86	0.000	20.16	24.69	R
	1	1.87	0.7981	9.14	12.53	FR
feeder_futures, feeder_index	0	94.59	0.000	20.16	24.69	R
	1	2.14	0.7493	9.14	12.53	FR
feeder_futures, NE_feeder_H	0	37.81	0.000	20.16	24.69	R
	1	1.82	0.8065	9.14	12.53	FR
TX_feeder_S, feeder_futures	0	55.56	0.000	20.16	24.69	R
	1	1.64	0.8393	9.14	12.53	FR
feeder_futures, KS_feeder_H	0	31.49	0.000	20.16	24.69	R
	1	1.95	0.783	9.14	12.53	FR
OK_feeder_S, MO_feeder_S	0	154.48	0.000	20.16	24.69	R
	1	1.84	0.8028	9.14	12.53	FR
KS_feeder_S, OK_feeder_S	0	158.68	0.000	20.16	24.69	R
	1	1.85	0.8023	9.14	12.53	FR
MT_feeder_S, OK_feeder_S	0	143.58	0.000	20.16	24.69	R
	1	1.8	0.8107	9.14	12.53	FR
CO_feeder_S, OK_feeder_S	0	61.67	0.000	20.16	24.69	R
	1	1.66	0.8359	9.14	12.53	FR
NE_feeder_S, SD_feeder_S	0	433.35	0.000	20.16	24.69	R
	1	1.76	0.8179	9.14	12.53	FR

Table 3. Johansen's Trace Tests, cont'd

	Johansen's Trace Tests for Cointegration					
price pair	Rank	LR	p-value	95%	99%	Decision
NE_feeder_S, CO_feeder_S	0	66.49	0.000	20.16	24.69	R
	1	1.53	0.8578	9.14	12.53	FR
CO_fed_S, CO_fed_H	0	186.45	0.000	20.16	24.69	R
	1	2.3	0.7184	9.14	12.53	FR
NE_fed_S, NE_fed_H	0	161.24	0.000	20.16	24.69	R
	1	2.22	0.7328	9.14	12.53	FR
TXOK_fed_S, TXOK_fed_H	0	110.69	0.000	20.16	24.69	R
	1	2.2	0.7381	9.14	12.53	FR
KS_fed_S, KS_fed_H	0	215.23	0.000	20.16	24.69	R
	1	2.27	0.725	9.14	12.53	FR
NE_fed_S, CO_fed_H	0	107.95	0.000	20.16	24.69	R
	1	2.23	0.7315	9.14	12.53	FR
KS_fed_H, TXOK_fed_H	0	170.95	0.000	20.16	24.69	R
	1	2.19	0.7394	9.14	12.53	FR
TXOK_fed_H, NE_fed_S	0	67.63	0.000	20.16	24.69	R
	1	2.18	0.7403	9.14	12.53	FR
IA_fed_S, IA_fed_H	0	121.2	0.000	20.16	24.69	R
	1	2.1	0.7564	9.14	12.53	FR
IA_fed_S, NE_fed_S	0	74.43	0.000	20.16	24.69	R
	1	2.17	0.7426	9.14	12.53	FR
fed_futures, NE_fed_S	0	59.05	0.000	20.16	24.69	R
	1	2.39	0.7023	9.14	12.53	FR
CO_feeder_S, feeder_futures	0	43.71	0.000	20.16	24.69	R
	1	1.58	0.8492	9.14	12.53	FR
fed_futures, feeder_futures	0	18.69	0.081	20.16	24.69	R
	1	1.93	0.7865	9.14	12.53	FR
box_beef, feeder_futures	0	22.17	0.025	20.16	24.69	R
	1	1.94	0.7846	9.14	12.53	FR

Appendix D Relative Adjustment Coefficients

Table 4. Speed of Adjustment and Relative Adjustment Coefficients

Price pair	Adjustment Coefficients		Relative Adj. Coefficients	
	α_1	α_2	θ_1	θ_2
MO_feeder_H, OK_feeder_H	0.435	0.142	0.246	0.754
KS_feeder_H, OK_feeder_H	0.226	0.287	0.559	0.441
MT_feeder_H, KS_feeder_H	0.230	0.132	0.365	0.635
CO_feeder_H, KS_feeder_H	0.181	0.100	0.356	0.644
TX_feeder_H, KS_feeder_H	0.369	0.155	0.296	0.704
NE_feeder_H, SD_feeder_H	0.223	0.388	0.635	0.365
feeder_futures, feeder_index	0.072	0.276	0.793	0.207
feeder_futures, NE_feeder_H	0.055	0.147	0.728	0.272
TX_feeder_S, feeder_futures	0.137	0.061	0.308	0.692
feeder_futures, KS_feeder_H	0.059	0.105	0.640	0.360
MO_feeder_S, OK_feeder_S	0.371	0.089	0.193	0.807
KS_feeder_S, OK_feeder_S	0.296	0.168	0.362	0.638
MT_feeder_S, OK_feeder_S	0.312	0.092	0.228	0.772
CO_feeder_S, OK_feeder_S	0.086	0.094	0.522	0.478
NE_feeder_S, SD_feeder_S	0.341	0.377	0.525	0.475
NE_feeder_S, CO_feeder_S	0.164	0.032	0.163	0.837
CO_fed_S, CO_fed_H	0.581	0.248	0.299	0.701
NE_fed_S, NE_fed_H	0.115	0.620	0.844	0.156
TXOK_fed_S, TXOK_fed_H	0.282	0.230	0.449	0.551
KS_fed_S, KS_fed_H	1.887	0.956	0.336	0.664
NE_fed_S, CO_fed_H	0.155	0.592	0.793	0.207
KS_fed_H, TXOK_fed_H	0.367	0.294	0.445	0.555
TXOK_fed_H, NE_fed_S	0.140	0.096	0.407	0.593
IA_fed_S, IA_fed_H	0.202	0.350	0.634	0.366
IA_fed_S, NE_fed_S	0.229	0.016	0.065	0.935
fed_futures, NE_fed_S	0.051	0.131	0.720	0.280
KS_feeder_S, feeder_futures	0.111	0.068	0.380	0.620
fed_futures, feeder_futures	0.031	0.012	0.279	0.721
box_beef, feeder_futures	0.038	0.010	0.208	0.792

Note: A high θ means low α and the market doesn't respond to the unpredicted shock in the market. It implies the market with a higher θ value is the price discovery. Winners of the tournament, i.e., price discovery is indicated in bold font.