CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

A Tournament Approach to Price Discovery in the U.S. Cattle Market

Jeffrey Wright[1], Man-Keun Kim[2,*] , Hernan A. Tejeda[3] and Hwa-Neyon Kim[4]

[1]Department of Agricultural Economics, Texas A&M University, College Station, TX, USA, [2]Department of Applied Economics, Utah State University, Logan, UT, USA, [3]Department of Agricultural Economics and Rural Sociology, University of Idaho, Moscow, ID, USA and [4]Department of Applied Economics, Jeju National University, Jeju City, Jeju, Korea
*Corresponding author. Email: mk.kim@usu.edu

Abstract
The dominant market where information is discovered plays the role of price leader providing substantial market information to other markets. This study investigates the dynamic relationships of 30 cattle markets across regions, cattle types, and cash/futures markets. The comparison of many markets, using an error correction model, is accomplished with the introduction of a tournament with a hierarchical cluster analysis, which allows us to conclude that the leading price for the U.S. cattle markets is discovered in the futures markets for both feeder and fed cattle.

1. Introduction

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 information that is relevant to other markets (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. This study identifies dynamic price characteristics and relationships by implementing a novel technique within a time series approach that may assist market participants along the cattle supply chain, specifically in procurement strategies, price risk mitigations, and potential investment decision-making for a number of businesses. Note that there have been a series of studies with time series approaches playing a larger role in price discovery, to identify a reference cattle market, as discussed below. One challenge is that time series approaches limit the number of markets to compare even though there may be many interrelated prices, in this case at least 30-price series. Prior studies have considered at most five to six markets. This paper attempts to overcome this challenge by combining time series models with a tournament method and using both in conjunction with the cluster analysis.

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)." 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 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 (1991) 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. One recent study, Coffey, Pendell, and Tonsor (2019), applied error correction model (ECM), directed acyclic graph, and Granger causality analysis to weekly negotiated live steer and heifer price data from June 2001 to July 2017, seeking to identify price discovery among negotiated cattle cash prices in the five major live cattle marketing regions. Interestingly, they found that Colorado has become the price leader, which is seemingly a smaller market. As results are mixed with respect to previous studies, more research may be warranted.

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. 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) ECM, are used to estimate price discovery weights. 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 has been used by Schwarz and Szakmary (1994), Foster (1996), Eun and Sabherwal (2003), Figuerola-Ferretti and Gonzalo (2010), and Arnade and Hoffman (2015) to identify price discovery among related markets.

However, this technique using the ECM with the relative ratio of the speed of adjustments is applicable only for the bivariate case, that is, 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 newly developed technique allows us to compare many cattle market prices across regions, cattle type, and cash/futures markets. The main objective of this study is to investigate the dynamic relationships between U.S. cattle markets, across regions, cattle types, and cash/futures markets. This study attempts to determine the reference U.S. cattle price using the tournament approach and cluster analysis. This study is structured as follows. After providing a general introduction to the price discovery process, we explore the concept of price discovery and introduce cluster analysis. We then present data and the empirical results. We finally draw conclusions of the study and outline future studies.

2. Price Discovery and Cluster Analysis

2.1. Price discovery

Price discovery is about determining which market is more informative. Ward and Schroeder (2002) define 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" (p. 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, 1995). The underlying idea behind price discovery 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 1 and they move together, that is, cointegrated. Suppose that, for some reasons, a shock
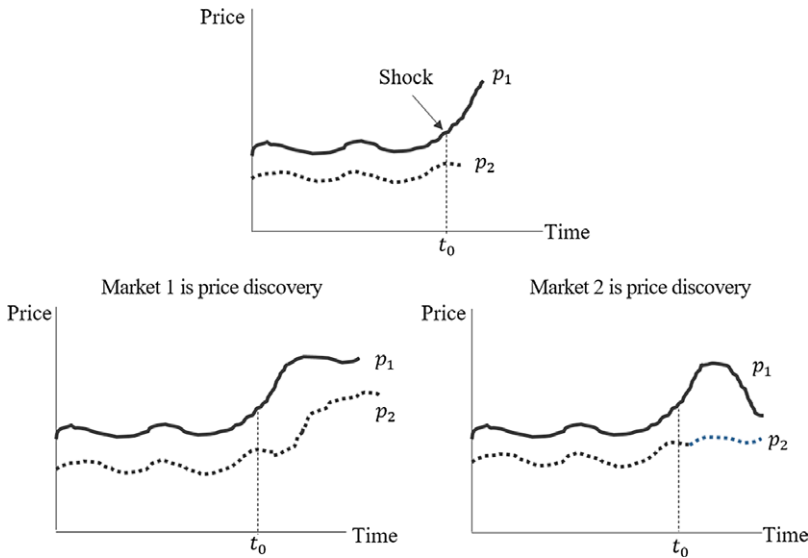
Figure 1. Price discovery illustration.

occurs in market 1, causing p_1 to rapidly increase (first panel, Figure 1).[1] A number of things can occur after this change. Let us assume that market 1 is the price leader. 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 1. 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 is more informative; market 2 (follower) adjusts its price more to close the gap when the price gap widens between markets 1 and 2. Let us now assume market 2 is the price leader. 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 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 1.

2.2. Vector error correction model and its limitation

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 has simply compared the speed of adjustment coefficients in the bivariate vector error correction model (VECM) as a share of the total adjustment, as developed in Gonzalo and Granger (1995) and expanded in Theissen (2002). Other measures of price discovery used in the literature from a multiple market perspective are information share (IS) of Hasbrouck (1995) and component share (CS) of Booth et al. (1999) and Chu et al. (1999). These are discussed at length in the *Journal of Financial Market*'s special issue on price discovery measurement in 2002 (Baillie et al., 2002; De Jong, 2002; Harris et al., 2002; Hasbrouck, 2002; Lehmann, 2002). The IS measures each market's relative contribution to the variance of the efficient price, while the CS

---

[1]Note that most shocks that occur in the cattle market are general market shocks rather than local market shocks as assumed here. If market 1 leads when there is a shock outside either market, then markets 1 and 2 respond but market 1 responds the least. Both markets may have idiosyncratic shocks.

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 of a VECM. Yan and Zivot (2010) proposed the joint use of the IS and the CS, and De Jong and Schotman (2010) developed the unobserved component approach.

The long-run equilibrium between two prices can be written as:

$$p_{1,t} = \beta_2 p_{2,t} + c + u_t \quad \Leftrightarrow \quad u_t = p_{1,t} - \beta_2 p_{2,t} - c \tag{1}$$

where $p_{1,t}$ and $p_{2,t}$ represent prices in markets 1 and 2 at time t, respectively. The term c (constant term) accounts for average differences in these two markets. The term u_t is the (long-run) error, which is zero in equilibrium.

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

$$\Delta p_{1,t} = \alpha_1 \big(p_{1,t-1} - \beta_2 p_{2,t-1} - c \big) + \sum_{k=1}^{K-1} \gamma_{11,k} \Delta p_{1,t-k} + \sum_{k=1}^{K-1} \gamma_{12,k} \Delta p_{2,t-k} + e_{1,t}$$

$$\Delta p_{2,t} = \alpha_2 \big(p_{1,t-1} - \beta_2 p_{2,t-1} - c \big) + \sum_{k=1}^{K-1} \gamma_{21,k} \Delta p_{1,t-k} + \sum_{k=1}^{K-1} \gamma_{22,k} \Delta p_{2,t-k} + e_{2,t} \tag{2}$$

where $\Delta p_{i,t}$ represents 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:

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

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 $\boldsymbol{\mu} = -\boldsymbol{\Pi}^{-1}\mathbf{c}$. K is the lag-length of the 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

$$\theta_1 = \frac{|\alpha_2|}{|\alpha_1| + |\alpha_2|}, \theta_2 = \frac{|\alpha_1|}{|\alpha_1| + |\alpha_2|}, \text{ and } \theta_1 + \theta_2 = 1 \tag{4}$$

A high θ indicates a low α (speed of adjustment toward a new long-run equilibrium), which implies that this market slowly responds to an unpredicted shock in the system; therefore, the market is the price discovery reference market. If $\theta_1 = \theta_2 = 0.5$, both markets contribute equally to the price discovery process, that is, both markets move at a similar rate toward the long-run equilibrium. Note that θ_i is applicable only for the bivariate case with $\boldsymbol{\alpha}_{2 \times 1}$. For the multivariate case with one cointegrating vector, we may arrange price discovery according to relative magnitude of α_i, but it is not clear how to calculate θ_i. Moreover, if there are multiple cointegrating vectors, comparing price discovery using α_i is not applicable.

However, an Achilles heel to price discovery using the speed of adjustment coefficients with the VECM outlined in equations (2) and (4) is its inability to compare more than two variables at a time. This is because there must be one cointegrating relationship between the relevant variables. For the multivariate case with one cointegrating relation, we may arrange price discovery according to relative magnitude of α_i but it is not clear how to calculate θ_i. Another shortcoming is that some markets may result in possibly having a negative weight generated from potential collinearity. In addition, in the multivariate case there might be more than one cointegrating relationship. 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.

2.3. Cluster analysis and tournament

As noted, when there are multiple markets, this technique, equation (4), may not be applicable. The tournament approach using cluster analysis provides a way to overcome this problem as demonstrated in Kim, Tejeda, and Wright (2016). 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 ECM to compare each series to every other series. This method may be inefficient and labor intensive, as growth in the number of markets being studied involves a substantial increase in the number of comparisons. An approximation to a round robin tournament approach was a study by Oellermann, Brorsen, and Farris (1989) which 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. Nonetheless, it was technically not a round robin tournament since possible combinations such as (5) live feeder futures vs. live cattle futures and (6) live cattle futures vs. live cattle cash were not included.

Another, more efficient, approach is to use a single elimination tournament. Our paper uses 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 loses 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 do not 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 cumbersome load it would entail, we will use a single elimination tournament 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 subcategories of the U.S. cattle market.

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 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 are 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. 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:

$$d(i,j) = \sqrt{\sum_{t=1}^{T} \left(p_{i,t} - p_{j,t} \right)^2} \tag{5}$$

Note that the Euclidean distance is nonnegative, the distance of a market to itself is 0, that is, $d(i,i) = 0$ and the distance is symmetric, that is, $d(i,j) = d(j,i)$. Equation (5) is converted to the root mean square deviation (RMSD) if the expression inside the square root is divided by T. A limitation of using equation (5) for the cattle market is of not accounting for different economic factors on how the market prices are determined. Prices may vary based on transportation costs, seasonality, product transformation, etc. (i.e. heterogeneous). These factors may create large differences between the prices of markets being considered as centers of price discovery. Thus, the differences in prices may reflect many potential economic factors that may not be price discovery related. Future work considers methods that account for these economic factors within the price equilibrium relationships.

Clustering comes in different types, including nested and unnested, that is, 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 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 2 shows an 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. A 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 2 shows a dendrogram for the five objects presented in Figure 2, where $l = 0$ shows the five objects as singleton clusters at level 0. At $l = 1$, objects a and b
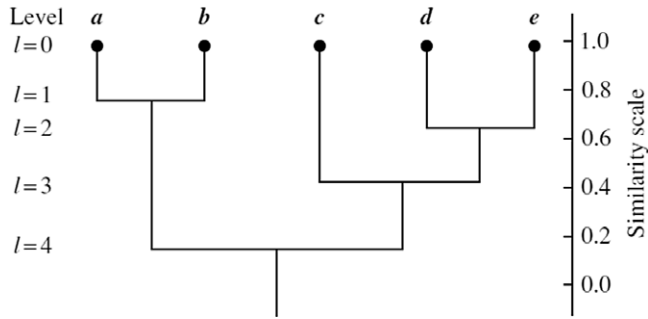
Figure 2. Dendrogram representation for hierarchical clustering. (Reproduced from fig. 10.6 in Han, Kamber, and Pei (2012), p. 460).
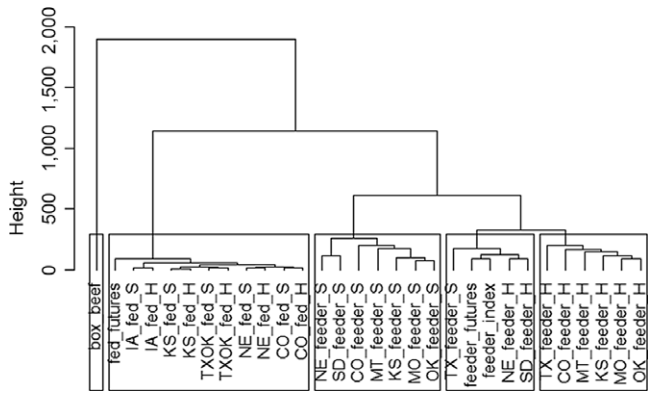

Figure 3. Cluster analysis dendrogram.
Notes: (1) State_fed_S = fed cattle steer in state; State_fed_H = fed cattle heifer in state; State Name_feeder_S = feeder cattle steer in state; State Name_feeder_H = feeder cattle heifer in state. (2) Vertical axis, Height, is the Euclidean distance between markets; height of 100 is approximately $3.5/cwt difference between markets (see equation (5)). (3) Boxes cut and visualize the dendrogram to five clusters (authors' choice).

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 3), they are merged together to form a single cluster.

In this research, the $\mathtt{hclust}$ function in R software is utilized. This program uses the maximum distance, $d_{max}(C_i, C_j) = \max_{p \in C_i, p' \in C_j} \{|p - p'|\}$, where $|p - p'|$ is the distance between two markets, to measure the distance between clusters. It is sometimes called a complete-linkage algorithm.

3. Data and Tests

3.1. Cattle price data

This study uses weekly cattle price data from May 2001 through October 2016. All cattle cash and futures price data are obtained from the Livestock Marketing Information Center (LMIC), which compiled the U.S. Department of Agriculture Agricultural Marketing Service (USDA-AMS) Livestock Mandatory Price Reporting (LMR) report data. The fed cattle price data were collected from all Free on Board (FOB) negotiated trade prices within a region. Regions studied include Texas–New Mexico–Oklahoma (TX–NM–OK), Kansas (KS), Nebraska (NE), Colorado (CO),

and Iowa–Minnesota (IA-MN). A weekly fed cattle price for each region represents the weighted average from the steers and heifers prices. All weights and grades of cattle were included in this calculation. For feeder cattle price, 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. A simple, unweighted average was calculated for both steers and heifers considering the specified weight ranges.[2]

Unfortunately, both the fed cattle price and the feeder cattle price 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 feeder cattle price data. Prices greater than $3/cwt were removed from the data, and these missing observations were estimated using linear interpolation. The price level of $3/cwt was chosen to maximize the amount of data retained, while still eliminating outliers and probable mistakes in the raw data.

Boxed beef cutout price is also included and it is provided by the weekly average beef cutout prices for Choice 600–900 lbs. from the "National Weekly Boxed Beef Cutout and Boxed Beef Cuts—Negotiated Sales" report by the Agricultural Marketing Service of the USDA and obtained from LMIC. The raw boxed beef cutout data had no missing observations and were used as provided. In addition to these prices, the feeder cattle index is included. The feeder cattle index is a 7-day weighted average of the total dollars sold divided by the total pounds during the same period. The feeder cattle index is included to gauge the level of the potential price discovery type of information that the index could possibly provide. The USDA-AMS issues daily reports which contain cattle eligible for the index. Cattle futures prices are collected from LMIC as well, which was compiled from Chicago Mercantile Exchange. The futures prices are of nearby contracts referring to the contracts' close price that is closest to expiration. It rolls over on the first date of the new month after the preceding contracts end. The weekly averages are calculated by taking the previous Monday–Friday for the week ending on a Friday data. In total, we have 30 cattle price series which are the combinations of regions, feeder/fed, steer/heifer, and futures prices. Table 1 presents basic statistics.

3.2. Stationary tests and cointegration tests

Before moving to cointegration tests to check if there exists one cointegrating relationship among pairs of series, stationarity tests for each price series were conducted. This study utilizes the Phillips Perron test to check for stationarity of each variable (Phillips and Perron, 1988). As expected, all the cattle prices are non-stationary (results are not reported to save space). All matched series were tested for cointegration using the Johansen's trace test, and all were found to be cointegrated. The fact that all these pairs are cointegrated makes them suitable for analysis with the ECM. The optimal lag of VECM is determined using the Schwarz information criterion. The number of cointegration relationship in the bivariate case is at most one cointegrating vector and can be investigated based on likelihood ratio (LR)-type tests (Lütkepohl and Krätzig, 2004). Johansen (1995) provides critical values for the LR test, which is known as the trace test.

---

[2]Averaging 500–900 pound price series accounted for all of mentioned weight categories. However, this generates limitations since aggregation from the average of these prices series for steers and heifers, per region, may result, for example, the 700 lb. price series had an average ±10% variation difference with respect to the 500 and 900 lb. price series, respectively. Moreover, the alternative of considering five different price series for heifers and five different price series for steers for each region increases dramatically the level of computation, not necessarily providing more insight into the findings.

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.

4. Results

4.1. Cluster analysis

The cluster analysis provides the cluster dendrogram as shown in Figure 3. 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
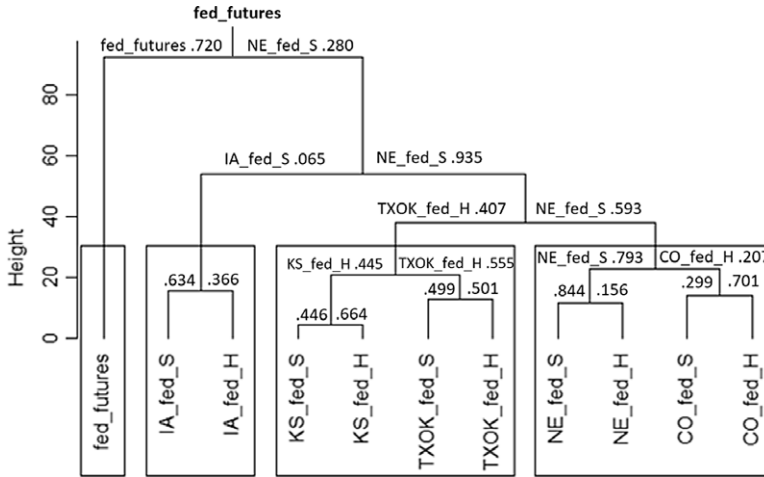
Figure 4. Fed cattle tournament result.
Notes: (1) State_fed_S = fed cattle steer in state; State_fed_H = fed cattle heifer in state. (2) Vertical axis, Height, is the Euclidean distance between markets; height of 100 is approximately $3.5/cwt difference between markets (see equation (7)). Blue boxes cut and visualize the dendrogram to four clusters (authors' choice). (3) Numbers are relative adjustment coefficients, that is, θ_i in equation (4); between two comparing markets, the winner has a larger θ.

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 hierarchy (second blue box in Figure 3). This 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 4. As previously explained in equation (5), this number represents the Euclidian distance between markets i and j. We can use this equation (5) to calculate the RMSDs, which is

$$\text{RMSD} \approx \sqrt{\frac{d(i,j)^2}{T}} \tag{6}$$

where T is the total number of observations.

We can now use 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 4 that the height on the box containing the fed cattle markets (second box in Figure 4) is approximately 100. We have 809 observations. So, the RMSD in price for the time periods examined for the two series furthest apart in the fed market is $3.52.[3] In contrast, the height for the feeder market is about 700. This gives the RMSD, within the feeder market of $24.60.

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 separate markets, rather than subcategories within the same market. The distance, for example, between the fed and feeder markets is approximately 1,200. This translates to the RMSD at each time period of $42.16. The distance between the boxed beef market and the other markets studied is about 2,000. This translates into an average price

---

[3]The height of the dendrogram is the distance, $d(i,j)$ from equation (5), and it could be converted to RMSD in equation (6), for example, $\sqrt{\frac{d(i,j)^2}{T}} = \sqrt{\frac{100^2}{809}} \approx 3.52$.

spread of $70.27. Although there are obvious connections between these markets, these large price spreads demonstrate the significant separations between these markets. This is at least partially due to boxed beef prices being in different units, as they are in price per wholesale pound rather than live pound. Due to these conditions, we identify 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 fed cattle price series as well as fed cattle futures (second blue box in Figure 3).
2. *Feeder cattle market*
   a. *Feeder steer market*: this market includes all feeder steers except those from Texas (third blue box in Figure 3).
   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 3).

We examine these submarkets and then compare the winners from each category. A VECM for each pair of these submarkets 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."

4.2. The tournament results of the fed cattle market

The first group we examine is the fed cattle market. Figure 4 summarizes the results. Figure 5 includes the name of the series being compared, followed by θ_i in equation (4) using the adjustment coefficients from the corresponding VECM. The results of relative adjustment coefficients are given in Table A1 in the appendix. 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 0.299, and the θ value for CO fed heifer is 0.701. 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. It is noteworthy that the dendrogram and winners of the fed cattle market in Figure 5 are similar to the size of the negotiated live cattle sales in recent years. In 2014–2018, NE had the largest share of 0–30-day negotiated cattle sales represented at 37% followed by IA at 21% and KS 17% (Schroeder, Schulz, and Tonsor, 2019). Earlier studies by Bailey and Brorsen (1985) found that TX was the center of price discovery; however, Schroeder and Goodwin (1990) identified Iowa–Nebraska as the leading price discovery locations; thus, changes in negotiated sales may have played a role.

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 fed cattle price markets are considered, that is, as shown in Figure 5, NE fed steer plays 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 the results from an extensive literature on futures markets going back at least to the early 1940s (Working, 1942).

4.3. The tournament results of feeder heifer and feeder futures markets

The second group includes the feeder heifers, feeder futures, the feeder index, and TX feeder steers (Figure 5). 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
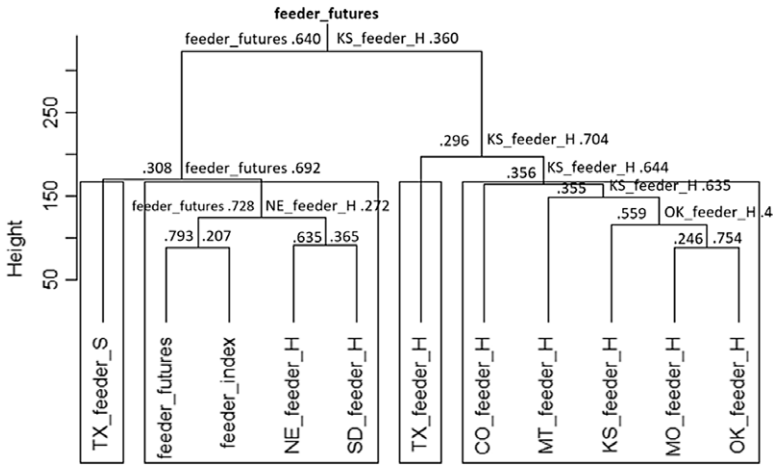
Figure 5. Feeder cattle heifer tournament result.
Notes: (1) State Name_feeder_S = feeder cattle steer in state; State Name_feeder_H = feeder cattle heifer in state. (2) Vertical axis, Height, is the Euclidean distance between markets; height of 100 is approximately $3.5/cwt difference between markets (see equation (7)). Blue boxes cut and visualize the dendrogram to four clusters (authors' choice). (3) Numbers are relative adjustment coefficients, that is, θ_i in equation (4); between two comparing markets, the winner has a larger θ.

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 heifer plays a prominent role in the price discovery process. This, however, would be a mistake. A closer examination of the bracket reveals KS feeder heifer 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.

4.4. The tournament results of feeder cattle steers

The third group includes all the feeder steers except those from TX. As the diagram in Figure 6 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 θ value for Colorado feeder steers to be 0.380 and the θ value for feeder futures to be 0.620. 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.

5. 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 study investigates the dynamic relationships of 30 cattle markets across regions, cattle types, and cash/futures markets. Determining price discovery process among a very large number of related markets is not efficiently workable. The comparison of many markets, using an ECM, is accomplished with the introduction of a tournament with a hierarchical cluster analysis, which allows us to conclude that the appropriate price for the U.S. cattle markets is discovered in the futures markets for both feeder and fed cattle.
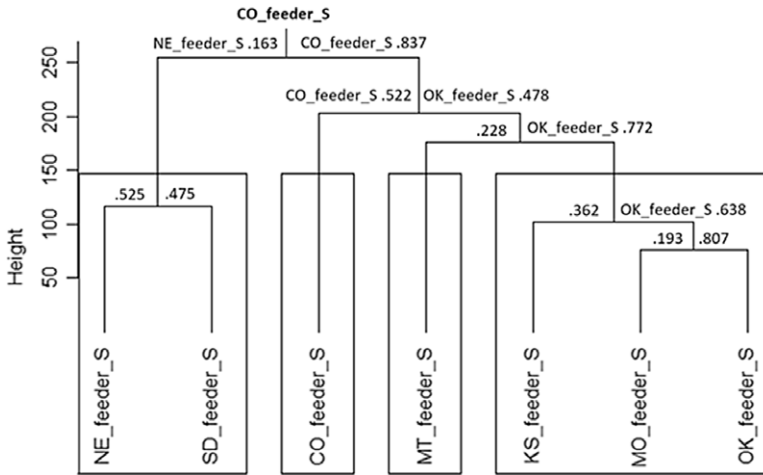
Figure 6. Feeder cattle steers tournament result.
Notes: (1) State Name_feeder_S = feeder cattle steer in state; State Name_feeder_H = feeder cattle heifer in state. (2) Vertical axis, Height, is the Euclidean distance between markets; height of 100 is approximately \$3.5/cwt difference between markets (see equation (5)). Blue boxes cut and visualize the dendrogram to 4 clusters (authors' choice). (3) Numbers are relative adjustment coefficients, that is, θ_i in equation (4); between two comparing markets, the winner has a larger θ.

The major contribution of this paper is the tournament approach with a cluster analysis.[4] 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 cumbersome load 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 ECM. The cluster analysis was conducted, and the resulting dendrogram was used to construct a single elimination tournament. The tournament was conducted using the estimated (bivariate) ECMs 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. It is noteworthy to mention that market weights and overall results were obtained using the distance function from equations (5) and (6), for example, steer and heifer prices being separated in different clusters. As noted earlier, these results may change if a different metric is used to measure the difference between the markets' equilibrium prices, for example, a method that incorporates economic factors that vary among the markets (e.g. transportations costs). The objective of this study was to provide a novel method to address a challenging dynamic price discovery process that has too many prices and to provide a way to simplify it. Limitations and future lines of study regarding cattle markets, or others, are below. The aim of this study was to provide a fair approximation of this process that has too many markets, while addressing the econometric challenge.

6. Limitations and Future Work

This study extends the literature by providing a method for the comparison of a large number of series in the price discovery process; however, there are limitations and still much work to be done.

---

[4]Kim, Tejeda, and Wright (2016) used a cluster and tournament approach to identify the price reference among U.S. milled rice markets. The present paper has evolved from Kim, Tejeda, and Wright (2016), including comments and feedback from conference attendees.

First, clustering cattle markets using the price Euclidean difference (or distance) may limit the segmentation of markets as there may be two prices which are relatively far apart in distance but potentially move together. Also, it is possible that prices differ based on transportation costs, seasonality, and product transformation (levels of supply chain) which are not price discovery related. In addition, aggregating price series using a simple average, even though it provides a good approximation, may mask the valid differences among the price series. Future study will focus on fewer cattle markets, specifically breaking up feeder calf prices. Moreover, use of feeder price index, an artificially created price series that overlaps with some of the cattle price data, will be part of this future study. Second, some of the markets could be equally important. In other words, values of θ in equation (4) are similar and not statistically significant. It is hard to decide which market is a winner for that case. To proceed the discussion we may decide a winner, ignoring the relative importance of the markets. However, the flexibility and continuous weighing of markets might be one of the advantages of the Gonzalo and Granger (1995) approach over Granger causality test which is discrete. Third, a tournament method in conjunction with cluster analysis is used in this paper to expand the Gonzalo and Granger (1995) approach, enabling the examination of multiple markets. In addition, this is not the only way to expand the Gonzalo and Granger (1995) approach. It suffers from the limitations inherent in single elimination tournament design. If the second strongest competitor is paired against the (resulting) 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. Using the ECM 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. In addition, it is obvious whether the losers in tournaments are responding to their cluster leader (winner) or there is a response to market leader (last winner of the tournament) beyond their own cluster leader.

References

Arnade, C., and L. Hoffman. "The Impact of Price Variability on Cash/Futures Market Relationships: Implications for Market Efficiency and Price Discovery." *Journal of Agricultural and Applied Economics* **47**(2015):539–59.

Bailey, D., and B.W. Brorsen. "Dynamics of Regional Fed Cattle Prices." *Western Journal of Agricultural Economics* **10**,1(1985):126–33.

Baillie, R.T., G.G. Booth, Y. Tse, and T. Zabotina. "Price Discovery and Common Factor Model." *Journal of Financial Markets* **5**,3(2002):309–22.

Booth, T., R. So, and Y. Tse. "Price Discovery in the German Equity Index Derivatives." *Journal of Futures Markets* **19**,6(1999):619–43.

Chu, Q., W. Hsieh, and Y. Tse. "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(1999):21–34.

Coffey, B.K., D.L. Pendell, and G.T. Tonsor. "Contemporaneous and Lagged Causal Relationships among Negotiated Live Cattle Markets." *Journal of Agricultural and Applied Economics* **51**,1(2019):182–98.

De Jong, F. "Measures of Contributions to Price Discovery: A Comparison." *Journal of Financial Markets* **5**,3(2002):323–7.

De Jong, F., and P.C. Schotman. "Price Discovery in Fragmented Markets." *Journal of Financial Econometrics* **8**(2010):1–28.

Eun, C., and S. Sabherwal. "Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks." *The Journal of Finance* **58**(2003):549–75.

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

Foster, A.J. "Price Discovery in Oil Markets: A Time Varying Analysis of the 1990-1991 Gulf Conflict." *Energy Economics* **18**(1996):231–46.

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

Han, J., M. Kamber, and J. Pei. *Data Mining Concepts and Techniques.* 3rd ed. Waltham, MA: Elsevier, 2012.

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

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

Hasbrouck, J. "Stalking the Efficient Price in Market Microstructure Specifications: An Overview." *Journal of Financial Markets* **5**,3(2002):329–39.

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

Kim, C. "Essays on Price Discovery." PhD Dissertation, Department of Economics, University of Washington, 2011.

Kim, M-K., H.A. Tejeda, and J. Wright. "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, 2016.

Koontz, S.R., P. Garcia, and M.A. Hudson. "Dominant-Satellite Relationships between Live Cattle and Cash Futures Markets." *The Journal of Futures Markets* **10**,2(1990):123–36.

Lehmann, B.N. "Some Desiderata for the Measurement of Price Discovery across Markets." *Journal of Financial Markets* **5**,3(2002):259–76.

Lütkepohl, H., and M. Krätzig. *Applied Time Series Econometrics.* Cambridge, UK: Cambridge University Press, 2004.

Oellermann, C.M., B.W. Brorsen, and P.L. Farris. "Price Discovery for Feeder Cattle." *Journal of Futures Markets* **9**(1989):113–21.

Phillips, P.C.B., and P. Perron. "Testing for a Unit Root in Time Series Regression." *Biometrika* **75**,2(1988):335–46.

Schroeder, T.C. "Spatial Fed Cattle Transaction Price Relationships." *Definition of Regional Cattle Procurement Markets*, M.L. Hayenga, S.R. Koontz, and T.C. Schroeder, eds. Washington, DC: U.S. Department of Agriculture, Grain Inspection, Packers and Stockyards Administration, GIPSA-RR 96-1, 1996.

Schroeder, T.C., and B.K. Goodwin. "Regional Fed Cattle Price Dynamics." *Western Journal of Agricultural Economics* **15**,1(1990):111–22.

Schroeder, T.C., and B.K. Goodwin. "Price Discovery and Cointegration for Live Hogs." *Journal of Futures Markets* **11**(1991):685–96.

Schroeder, T.C., L.L. Schulz, and G.T. Tonsor. *Feasibility Assessment of Reporting Negotiated Slaughter Cattle Purchases in Separate Delivery Window Categories.* Kansas City, MO: United States Department of Agriculture, Agricultural Marketing Service, 2019.

Schwarz, T.V., and A.C. Szakmary. "Price Discovery in Petroleum Markets: Arbitrage, Cointegration, and the Time Interval of Analysis." *Journal of Futures Markets* **14**(1994):147–67.

Tan, P-N., M. Steinbach, and V. Kumar. *Introduction to Data Mining*, 1st ed. London, UK: Pearson, 2006.

Theissen, E. "Price Discovery in Floor and Screen Trading Systems." *Journal of Empirical Finance* **9**(2002):455–74.

Ward, C.E. and T.C. Schroeder. "Price Determination versus Price Discovery." *Managing Today's Cattle Market and Beyond*, 2002. Internet site: http://marketing.uwagec.org/MngTCMkt/#update (Accessed October 26, 2020).

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

Yan, B., and E. Zivot. "A Structural Analysis of Price Discovery Measures." *Journal of Financial Markets* **13**(2010):1–19.

Appendix

Table A1. 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 does not 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, that is, price discovery, are indicated in bold font.