Journal of Agricultural and Applied Economics (2025), 1–19
doi:10.1017/aae.2025.10009

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Unraveling Hidden Patterns in Fed Cattle Negotiated Cash Prices Using Machine Learning

Zuyi Wang[1], Man-Keun Kim[2] and Hernan Tejeda[3]

[1]Department of Agricultural Sciences, Clemson University, Clemson, SC, USA, [2]Department of Applied Economics, Utah State University, Logan, UT, USA and [3]Department of Agricultural Economics and Rural Sociology, University of Idaho, Moscow, ID, USA
Corresponding author: Zuyi Wang; Email: zuyiw@clemson.edu

Abstract
The decline in fed cattle cash sales and its impact on price discovery are concerning. This study extends existing literature by utilizing machine learning to explore factors, particularly decision trees and random forests, to explore factors influencing fed cattle price ranges, complementing traditional regression analyses. These models uncover hidden patterns and provide additional insights into the cattle market. Key variables such as weight range, head count, and trade location, are found to be associated with price ranges. Notably, the weight range emerges as the primary variable influencing the price range, with smaller weight ranges linked to lower price ranges.

Keywords: Cattle markets; decision tree; price range; Random Forest

JEL classifications: C45; Q13; Q18

1. Introduction

The Livestock Mandatory Price Reporting (LMR) Act of 1999 requires packers to report transactional data twice a day to the United States (US) Department of Agriculture (USDA). The USDA then freely publishes these data for market participants to make informed decisions. The packers' identity and their transactions are confidential if certain criteria are not met.[1] The thinning sales and the reporting restrictions that can lead to transactions not being reported has motivated market participants to examine the accuracy of the negotiated cash sales. The LMR data includes the daily weighted average price, the daily minimum price, and the daily maximum price. However, market participants do not have access to the complete fed cattle price distribution,[2] but only certain price intervals and empirical observations of formula and negotiated cash sales. Consequently, sales reported of very minimum or very maximum prices and of values close to these rather extreme corners may have a disproportionate influence on the weighted average price. This influence can distort the resulting weighted average as a representation of the market's central tendency measure, particularly in case where the market distribution is significantly skewed.

Recent studies, such as Boyer et al. (2023), have explored the challenges and problems related to price discovery in the context of fed cattle negotiated prices, highlighting the implications of wide

[1]More on the "3/70/20" guidelines followed in LMPR, which restricts data aggregation, can be found at USDA AMS (2021).
[2]It is important to note that as of August 2021, weekly price distribution data is now available through LM_CT-215, the National Weekly Direct Beef Type Price Distribution report, accessible via DataMart, USDA AMS.

price ranges. Wide price ranges[3] may lead to significant price volatility and increasing the risk for both sellers and buyers. A range of policy proposals, including the 50/14 and Spot Market bills, as well as the Cattle Market Transparency Act of 2021, have been introduced to bolster negotiated cash sales within the U.S. cattle market and decrease fed cattle price ranges. However, the potential impact of these policy measures remains elusive and challenging to quantify. Boyer et al. (2023) conducted a comprehensive examination in pursuit of identifying factors influencing price ranges in the negotiated cash market, considering various aspects, such as the volume of head sold, the day of the week, sex, grade, weight range, and other pertinent factors, shedding light in their effects on price ranges within the negotiated cash market. Their findings uncovered intriguing patterns, such as price ranges peaking on Mondays and being lowest on Tuesdays, with a gradual increase from Wednesday to Friday. Moreover, the study revealed that price ranges increased with the volume of trade until reaching an approximate threshold of 8,800 head per sale per day, beyond which they began to taper off. Notably, the Iowa/Minnesota market demonstrated the highest negotiated cash price trade region relative to other regions.

Building on the work of Boyer et al. (2023), this study seeks to provide "additional" insights into the variability of fed cattle price ranges by employing machine learning techniques, particularly decision trees. Storm et al. (2020), highlighted limitations in regression models, noting their reliance on restrictive functional forms with limited theoretical grounding, which can result in challenges when attempting to estimate equations. Economic theory may not always offer clear guidance on the equations that are to be estimated. In addition, Storm et al. (2020) emphasized that econometric modeling may struggle to effectively handle a large number of right-hand side variables, which becomes a concern when researchers must contend with numerous potential exploratory variables. This situation becomes particularly critical when the causal relationships among variables are unclear, and when the variables themselves may exhibit high levels of multicollinearity. Additionally, it is worth noting that traditional regression approaches may not uncover information or patterns within the data that were not previously known. These approaches may fall short in addressing such matters, as researchers typically determine the variables included in the regression models based on their gained knowledge. These limitations inherent in traditional regression approaches serve as motivation for exploring alternative methods in analyzing the dynamics of price ranges in the fed cattle market, as discussed in this paper.

The primary objective of this study is to analyze the factors affecting daily price ranges of fed cattle negotiated cash sales using decision trees. Given the limitations of single decision trees, including their susceptibility to overfitting and sensitivity to minor data variations, our study incorporates ensemble methods, Random Forests (RFs), to improve the accuracy and stability by aggregating predictions from multiple trees. Our approach seeks to complement and extend the results of Boyer et al. (2023) by identifying similarities and new findings. Unlike Boyer et al. (2023), our paper identifies a new finding: weight range proves to be more influential in cattle markets than previously recognized. Moreover, our analysis also uncovered approximately 41.6% of reported transactions had a zero weight range coupled with a zero price range, a nuance not mentioned or considered in previous studies. While removal of this latter data did not significantly affect our estimated results, it does raise the question of the purpose it may serve and its benefit to market stakeholders.

Our findings have relevant implications for market participants and policy. By identifying key factors that contribute to price range disparities, this study informs market participants of which locations and selling basis, in addition to weight range, are anticipated to mostly affect the price range of cattle. As for policy, these findings may be conducive to examining the purpose of daily reports that indicate zero weight ranges coupled with zero price ranges, and the benefit in their

[3]Considering transactions that occur at different geographical locations, days of the week, volumes of trade, weight ranges, grade, transaction types and other (see Table 2 for table of characteristics) .

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

recording and dissemination. In sum, this research contributes to an enhanced understanding of the underlying factors affecting price ranges in negotiated fed cattle transactions, offering insightful information to market stakeholders and policy makers.

2. Market background and related studies

In the mid-2000s, the fed cattle market experienced a shift as formula pricing, negotiated grid pricing, and forward contract pricing began to replace negotiated cash sales, leading to the dominance of alternative marketing agreements (Adjemian et al., 2016). Among these, formula pricing became the most widely used AMA, setting prices based on negotiated cash prices while incorporating quality adjustments (Adjemian et al., 2016).

Thin markets, characterized by low trade volume, lead to diminished market information and increased price volatility. Existing literature on the impact of Livestock Mandatory Reporting (LMR) on fed cattle prices suggests that greater public information available to packers could potentially decrease price variability (Anderson et al., 2019). Azzam (2003) asserts that LMR enhances competition among packers, contributing to a decrease in price variance. However, the thinning of markets is not uniform across reporting regions, with variations in negotiated cash sales affecting the efficacy of hedging strategies and giving rise to regional price disparities (Pendell and Schroeder, 2006; Schroeder et al., 2019). These regional discrepancies in cash prices influence the informativeness of futures market price discovery and contribute to heightened price volatility (Schroeder et al., 2019).

Furthermore, the thinning of negotiated cash sales is also observed across different days of the week, as highlighted in various studies (Crespi and Sexton, 2004; Schroeder et al., 1993; Schroeter and Azzam, 2003; Ward et al., 1998; Ward, 1992). Research suggests that cattle markets tend to be subdued early in the week and gain momentum as the week progresses (Crespi and Sexton, 2004). Conversely, some studies indicate that most purchases occur on Mondays, with fewer sales observed later in the week. These studies also note that prices tend to be higher on days with increased trade activity (Schroeder et al., 1993; Ward et al., 1998; Ward, 1992). Despite these variations, scholars acknowledge the uneven distribution of cattle trade throughout the week, suggesting the possibility of a thinner fed cattle market on specific days. This uneven distribution can impact cattle feeders who are price takers and have limited control over the timing of their sales.

This study extends the existing literature by employing complementary quantitative analysis through machine learning techniques, specifically decision tree learners, to analyze negotiated cattle price data. The primary objective is to uncover latent patterns within the dataset and enhance the descriptive capabilities of the model. By addressing the inherent limitations associated with traditional regression methods, this approach aims to provide a deeper understanding of the dynamics in the fed cattle market. The goal is to offer stakeholders, including cattle feeders, packers, and policymakers, more comprehensive market information.

3. Decision trees and random forest

Decision tree learners are classification algorithms that employ a tree-like structure to model the relationships between features (variables) and possible outcomes (Lantz, 2013). The primary goal of partitioning is to minimize dissimilarity in the terminal nodes, that is, groups with similar response values (Boehmke and Greenwell, 2020). While various methodologies exist for constructing decision trees, the most well-known is the classification and regression trees proposed in Breiman et al. (1983). Decision tree is a supervised learning approach and its algorithm starts by identifying the single variable that can best divide the data into two distinct groups (Chiu, 2015). After identifying this variable, the data is partitioned accordingly, and this process is recursively applied to each resulting subgroup (Therneau et al., 2022). This recursive

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

partitioning continues until the subgroups either reach a predefined minimum size or further splits no longer yield significant improvements (Therneau et al., 2022).

Suppose that data comes in records of the form $(\mathbf{X}, \mathbf{y})$ where the dependent variable, $\mathbf{y}$, is the target variable that we are trying to understand, classify or generalize, in this case, fed cattle negotiated cash price range. The features or independent variables, $\mathbf{X}$, are denoted as $x_1, x_2, \cdots, x_k$, which include factors such as types of cattle, day of trading, regions, cattle weight, weight range etc. The objective at each node is to identify the optimal feature, x_i, to partition the data into two regions, R_1 and R_2, minimizing the overall error between the actual response y_i, and the predicted constant, c_i. (Boehmke and Greenwell, 2020; Rokach and Maimon, 2014). For regression problems[4], the objective function to minimize is the sum of squared errors (SSE) as defined below:

$$\text{SSE} = \sum\nolimits_{i \in R_1} -(y_i - c_1)^2 + \sum\nolimits_{i \in R_2} -(y_i - c_2)^2 \tag{1}$$

where SSE measures the heterogeneity in the within-node (Rokach and Maimon, 2014). Once the optimal feature/split combination is identified, the dataset is divided into two distinct regions, and this splitting process is iteratively applied to each of these regions. This recursive procedure persists until a predefined stopping criterion is met, such as reaching a maximum depth or when the tree becomes excessively intricate, as discussed by Boehmke and Greenwell (2020). It is worth highlighting that a single factor can be utilized multiple times within the tree structure.

Figure 1 illustrates a regression tree with a single root node, using simulated data for cattle price range and weight dispersion. The tree diagram shows that weight dispersion, denoted as wd, serves as the splitting criterion, with a split occurring at wd equal 9, where the SSE in equation (1) is minimized. This decision results in a single split based on the weight dispersion, as depicted on the right side of Figure 1. The resulting decision boundary indicates that when wd less than 9, the predicted value is \$0.67 per hundredweight (cwt) (33% of observations), while for wd greater than 9, it is \$2.5/cwt (67% of observations).

Figure 2 shows a "deeper tree" which continues to split based on wd, the only relevant feature in the example, until a pre-determined stopping criteria is met. The decision tree in Figure 2, with depth $= 2$, results in two decision splits along wd values and three prediction regions (right). The resulting decision boundary is shown on the left.

The challenge with decision trees is their tendency to develop complex structures, known as overfitting, which can result in poor generalization (Boehmke and Greenwell, 2020; Lantz, 2013). Consequently, we need to manage the depth and complexity of the tree to optimize predictive performance. To achieve this balance, researchers often employ pruning (Alpaydin, 2014; Boehmke and Greenwell, 2020; Rokach and Maimon, 2014), which involves developing a complex tree and then pruning it back to find an optimal subtree. The optimal subtree is determined by pruning the complex tree using a cost complexity parameter (cp), α, which penalizes the objective function in equation (1) for the number of terminal nodes of the tree, T:

$$\min \text{SSE} + \alpha|\text{T}| \tag{2}$$

For a given value of α, the goal is to identify the smallest pruned tree that minimizes the penalized error, similar to the lasso (least absolute shrinkage and selection operator) penalty (Tibshirani, 1996). As with regularization techniques, smaller penalties often produce more complex models, resulting in larger trees, while larger penalties yield much smaller trees. As a tree grows in size, the reduction in SSE must outweigh the cost complexity penalty. To pinpoint the optimal value of α and its resulting optimal subtree that generalizes most effectively to the unseen data, it is of common practice to evaluate multiple models across a range of α values and employ cross-validation.

[4]For classification tasks involving a categorical variable y_i, the partitioning strategy often aims to maximize the reduction in impurity measures, such as cross-entropy or the Gini index, as discussed by Boehmke and Greenwell (2020) and Rokach and Maimon (2014).

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

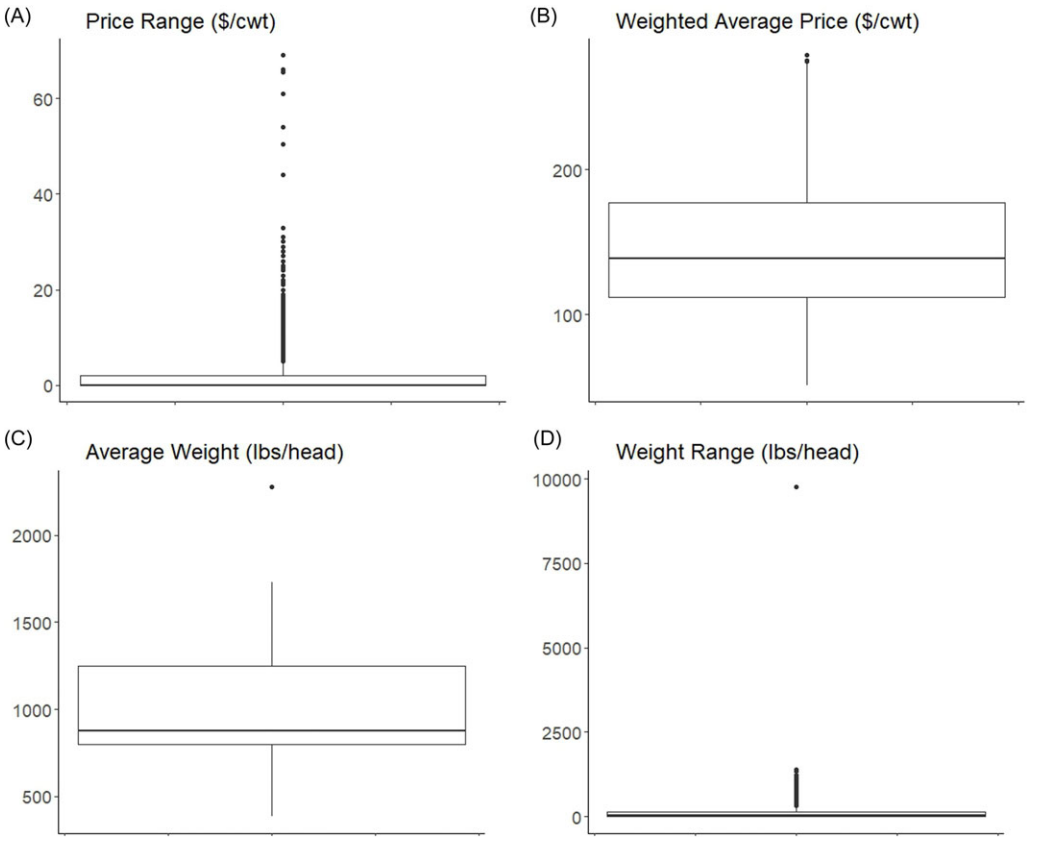

Figure 1. Box plots of cattle price range, average price, weight, and weight range. Box plots are created using the raw dataset with 138,956 observations. Dark circles represent potential outliers, defined as values exceeding Q3 + 1.5 IQR. However, note that price range and weight range are highly skewed with long right tails, making the IQR method less effective for detecting outliers in these variables.

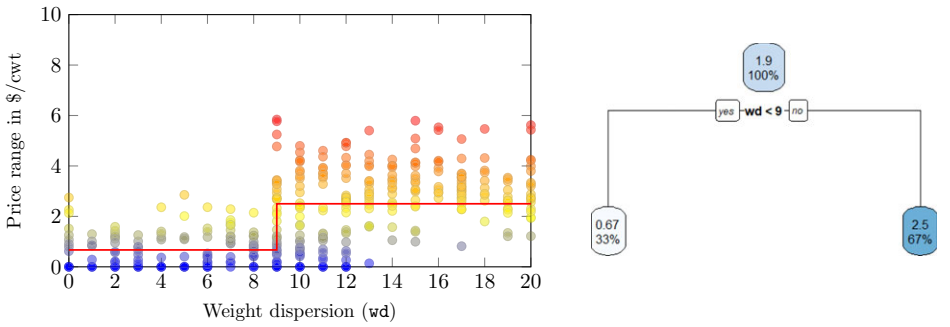

Figure 2. Decision trees demonstration using simulated price range and weight dispersion.

To ensure robust model performance and generalization, the dataset is divided into training and testing sets using an 80-20 split. Specifically, 80% of the data is used for training the decision tree model, allowing it to learn patterns and relationships between the input features (such as weight range, head count, and trade location) and the target variable (price range). The remaining 20% of the data serves as the testing set, which is reserved for evaluating the model's performance. This split helps in assessing how well the model generalizes to new, unseen data and ensures that

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

the results are not overly fitted to the training set. The training process involves constructing the decision tree by recursively partitioning the data to maximize the separation of price ranges, while the testing phase evaluates the model's accuracy and effectiveness in predicting price ranges based on the held-out data. This methodology enhances the reliability of the model's predictions and provides a comprehensive understanding of the factors influencing cattle price ranges.

RF (Breiman, 2001; Liaw and Wiener, 2002) is an ensemble learning method designed to enhance prediction accuracy and reduce overfitting by combining multiple decision trees. It constructs a "forest" of decision trees, where each tree is built using a random subset of the data and a random subset of features. The predictions from these trees are then aggregated, a process known as Bootstrap Aggregating (or "bagging"). This involves averaging the predictions for regression tasks or taking a majority vote for classification tasks. By training each tree on different samples of the data and using a subset of features for each split, RF reduces variance and improves overall model performance (Hastie et al., 2009). In the context of our study on cattle price ranges, RF offers significant advantages. It effectively manages a large number of potential explanatory variables and captures complex interactions among them. By combining the predictions from various decision trees, RF provides more accurate and stable results compared to individual decision trees. Additionally, RF can handle both numerical and categorical variables without requiring extensive data preprocessing (Fan et al., 2021), making it well-suited for analyzing survey data with diverse predictors. However, RF also has some drawbacks. The model's complexity can make it more difficult to interpret compared to simpler models like single decision trees (Ziegler and König, 2014). Additionally, RF can be computationally intensive, requiring more memory and processing power, particularly with a large number of trees or a large dataset. This increased computational cost may limit its application in environments with limited resources.

4. Data

To demonstrate the value of these innovative methods, this paper utilizes the data from Boyer et al. (2023), compiled from AMS which contains daily negotiated cash purchase data of fed cattle in Texas/Oklahoma/New Mexico, Kansas, Nebraska, Iowa, and Minnesota. The weekly data covers the years 2001 to 2019 and is categorized into three groups: (i) cattle-related variables, including daily negotiated cash price, cattle weight, head count, and cattle types in transactions; (ii) transaction-related variables, such as the date of the transaction and the month and year; and (iii) geographical variables, such as states.

To strengthen our analysis, we made box plots for key variables, including price range, weighted average price, average weight, and weight range, to identify potential outliers. While average weight values met expectations, both price range and weight range exhibited numerous outliers (see Figure 3). As shown, outliers were most pronounced in price (Panel A) and weight range (Panel D). Given the highly skewed distributions with long right tails, the traditional interquartile range (IQR) method was insufficient for detecting outliers. To address this, we used the 99th percentile as a threshold for identifying extreme values and ensure more accurate outlier detection. This method effectively captures the upper extremes of the distribution, accounting for the severe skewness in the data, which may due to error in measurement, mis-reporting or other. As a result, we excluded data where the price range exceeded $9/cwt and where the weight range was greater than 493 pounds per head. This adjustment reduced the dataset to 135,836 observations, dropping 3,120 data points.

Table 1 presents a summary of the cattle-related variables, including weighted average price, price range, headcount, and weight statistics across different selling terms. The weighted average price of cattle during this period was $144/cwt, with a minimum of $51/cwt and a maximum of $279/cwt. This variation in price reflects the different transaction types, as the prices for dressed delivered cattle averaged $169/cwt, dressed FOB cattle averaged $185/cwt, while live delivered and

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

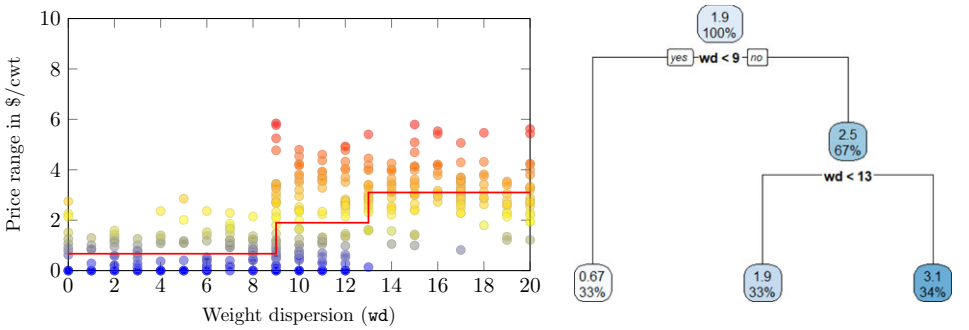

Figure 3. Decision trees demonstration using simulated price range and weight dispersion.

live FOB cattle prices were lower at $118/cwt and $106/cwt, respectively. The variation in prices across the selling terms suggests that the transaction type has a significant impact on the pricing of cattle. Dressed FOB transactions had the highest average price, followed by dressed delivered cattle, while live cattle transactions, especially live FOB, had lower average prices.

The price range, defined as the difference between the daily high price and the daily low price (Boyer et al., 2023), averaged $1.01/cwt, with a wide range up to $8.80/cwt. This indicates considerable variability in the price range. A detailed breakdown shows that the price range for dressed delivered cattle had an average of $1.11/cwt, quite higher than the price range for dressed FOB ($0.46/cwt) and slightly higher than live delivered ($0.98/cwt) cattle. The price range for live FOB cattle ($0.87/cwt) was also lower than both dressed sale terms. Figure 4 illustrates the distribution of price ranges. Approximately 58% of observations fall into zero differences (range) in prices, highlighting a significant concentration of data points at this value. The distribution exhibits a strong right skew with a long tail, reflecting a prevalence of low price ranges and few observations with much higher values. Notably, about 95% of observations have a price range of less than or equal to $5/cwt indicating that most price fluctuations are relatively small.

In terms of transaction characteristics, the average number of cattle per sale was 815 head, with the maximum number reaching 35,980 head. This variation highlights the scale and spread in the size of cattle sales. The average weight per head during the sample period was 1,002 pounds, with a broad range from 385 to 1,729 pounds, illustrating a large mix in cattle sizes across transactions. When comparing across the selling bases, live delivered and live FOB cattle had the highest average weight per head at 1,364 pounds and 1,283 pounds, respectively. In contrast, dressed delivered cattle had an average weight of 816 pounds, and dressed FOB cattle averaged 897 pounds. These differences show that live cattle, sold before slaughter, are larger as anticipated. Therefore, a $1 per cwt price range per head has a more significant impact on live cattle, as they are heavier.

Additionally, the weight range per head sold had an average of 76 pounds, with a wide range extending up to 492 pounds, showing considerable variation in cattle weights as well. Live cattle, especially live FOB, exhibited the largest weight ranges, with values of up to 492 pounds. This can be attributed to the broader spectrum of cattle sizes being sold in these markets. In contrast, dressed delivered and dressed FOB cattle had narrower weight ranges, reflecting a more homogeneous resulting processed cattle for these types of transactions.

Table 2 reports cattle transaction-related variables, including class, selling terms, grade, and locations of transaction data. Cattle are classified by breed and gender into dairy and beef types, which include steers, heifers, or mixed steer/heifer lots. Steers were the most predominant, comprising 33% of the total, followed by mixed steer/heifer lots at 29%, heifers at 27%, and dairy cattle at 11%. The selling terms for these transactions fall into two categories: dressed delivered or live free on board (FOB). Dressed refers to the carcass weight after slaughter, which is used for

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

Table 1. Basic statistics of cattle price variables from 2001 to 2019

Variables	Count	Mean	StDev	Min	Max
Price range[a] ($/cwt)	135,836	1.01	1.62	0	8.80
Dressed delivered	81,382	1.11	1.81	0	8.80
Dressed FOB	494	0.46	1.06	0	8.00
Live delivered	696	0.98	1.42	0	7.25
Live FOB	53,264	0.87	1.26	0	8.50
Weighted average price ($/cwt)	135,836	143.95	44.47	51	279
Dressed delivered	81,382	169.04	36.27	74	279
Dressed FOB	494	185.25	12.71	157	208
Live delivered	696	117.55	7.53	97	133
Live FOB	53,264	105.57	23.88	51	251
Number of cattle per sale	135,836	815	1,975	1	35,980
Dressed delivered	81,382	542	1,237	1	27,826
Dressed FOB	494	308	342	11	2,417
Live delivered	696	354	434	11	2,430
Live FOB	53,264	1,244	2,701	4	35,980
Average weight per head (lbs)	135,836	1,002	242.5	385	1,729
Dressed delivered	81,382	816	64.6	385	1,440
Dressed FOB	494	897	60.2	698	1,052
Live delivered	696	1,364	89.0	896	1,600
Live FOB	53,264	1,283	99.6	540	1,729
Weight range per head sold[b] (lbs)	135,836	75.9	98.8	0	492
Dressed delivered	81,382	56.7	83.6	0	492
Dressed FOB	494	23.5	37.4	0	216
Live delivered	696	49.7	68.6	0	325
Live FOB	53,264	106.0	112.4	0	491

[a]Price range is defined as the difference between the daily high price and the daily low price (Boyer et al., 2023).
[b]Weight range is defined as the difference between the highest average weight and the lowest average weight (Boyer et al., 2023).

payment calculation. Live FOB indicates cattle purchased while alive, with the buyer covering transportation costs. Most transactions involved dressed delivered sales, comprising 60%, followed by FOB at 39%, with dressed FOB and live delivered accounting for much smaller shares (0.04% and 0.05%, respectively).

Regarding location, trades were most common in Nebraska (38.5%), followed by Iowa/Minnesota (22.7%), Kansas (20.6%), and Texas/Oklahoma/New Mexico (18.2%). Note that these percentages reflect the distribution of trades, not the total number of cattle sold at each location. The grade variable indicates the proportion of each lot with a choice-grade quality designation after assessment. Lots are categorized into various grade ranges, including 0–35% choice, 35–65% choice, 65–80% choice, or over 80% choice. The distribution of these grades across the 35% to over 80% choice categories was fairly uniform, with 0–35% choice lots accounting for 7% of all trades.

Table 3 provides information on the timing of transactions, divided into two sections: Day of the week and Month of the year. Each section shows the total number of transactions for different

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

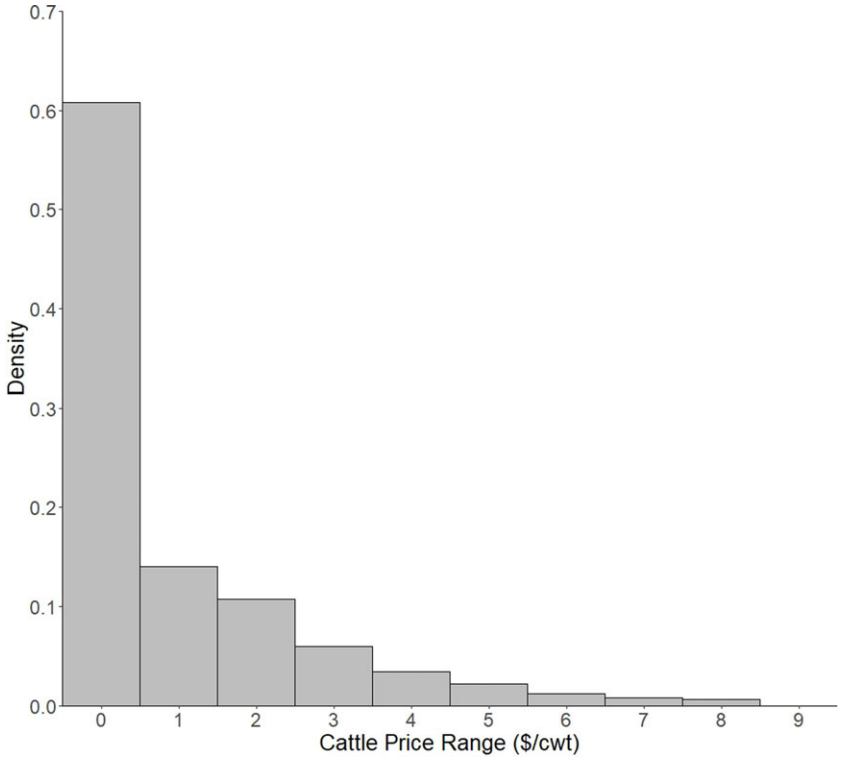

Figure 4. Cattle price range histogram.

Table 2. Basic statistics of cattle transaction variables from 2001 to 2019 ($n = 135,836$)

Class	Dairy cattle	Heifer	Mixed steer/heifer	Steer
	15,192	36,540	39,562	44,542
	(11.2%)	(26.9%)	(29.1%)	(32.8%)
Selling term	Dressed delivered	Dressed FOB	Live delivered	Live FOB
	81,382	494	696	53,264
	(60.0%)	(0.04%)	(0.05%)	(39.2%)
Location	IA/MN	KS	NE	TX/OK/NM
	30,880	27,916	52,291	24,749
	(22.7%)	(20.6%)	(38.5%)	(18.2%)
Grade	0–35% choice	35–65% choice	65–80% choice.	over 80% choice
	9,008	46,156	45,457	38,215
	(6.6%)	(31.8%)	(33.5%)	(28.1%)

days of the week or months of the year, along with the corresponding percentages. Table 3 offers insights into when transactions occur, both in terms of the day of the week and the month of the year. From the day of the week section, we observe that Mondays account for the highest number of transactions, with a total of 46,999 transactions, which represents 34.6% of the total transactions. This suggests that the beginning of the workweek is a popular time for sales. In

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

Table 3. Transaction count by day of week and month (2001–2019, $n = 135{,}836$)

Day of week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	46,999	9,420	14,848	26,542	38,026	1
	(34.6%)	(6.9%)	(10.9%)	(19.5%)	(28.0%)	(0.0%)
Month of year	Jan	Feb	Mar	Apr	May	Jun
	10,523	9,363	11,934	11,717	12,415	11,859
	(7.7%)	(6.9%)	(8.8%)	(8.6%)	(9.1%)	(8.7%)
	Jul	Aug	Sep	Oct	Nov	Dec
	11,530	12,326	11,398	11,400	10,625	10,746
	(8.5%)	(9.1%)	(8.4%)	(8.4%)	(7.8%)	(7.9%)

contrast, Tuesdays and Wednesdays see fewer transactions, with only 9,420 (6.9%) and 14,848 (10.9%) transactions, respectively. This indicates that mid-week days are less favored for sales. Table 3 also reveals a relatively stable pattern of sales throughout the year, with only minor variations. Notably, there is a slight increase in transactions during the summer months, such as May and June, while transaction counts are lower in winter, particularly in February. These patterns suggest a seasonal influence on transaction volumes, which may be anticipated.

5. Results and discussion

This section presents the findings from the analysis of cattle price ranges using decision tree and RF models. We begin by discussing the decision tree analysis, highlighting how the model interprets and splits the data based on different attributes. Following this, we assess the model's performance through validation metrics and compare it to the more complex RF model to understand the added value of ensemble learning methods.

5.1. Discussion of decision tree analysis

The results[5] with the default value of complexity parameter (cp) of 0.01 are presented in Figure 5. For building the decision tree, the dataset was divided into training and testing sets using an 80-20 split, where 80% of the data were used for training and 20% for testing. At the root node of the decision tree in Figure 5, the average price range is \$1.01/cwt, indicating that in the absence of additional information about the transaction, the expected price range is \$1.01/cwt. The first attribute used to split the dataset is the weight range. When the weight range is less than 1 lb/head, 41.6% of the dataset falls into a very small price range of 0.075/cwt. When the weight ranges are greater than or equal to 1 lb/head, the average price range increases to \$1.68/cwt, covering 58.4% of the data.

Further partitioning is based on the location. Further partitioning is based on the location. For the location attribute, transactions in Kansas, Texas/Oklahoma/New Mexico are further analyzed. When the location is within Kansas, Texas/Oklahoma/New Mexico, the price range decreases to \$1.22/cwt, affecting 21.5% of the dataset. If the location is not Kansas, Texas/Oklahoma/New Mexico, the price range rises to \$1.95/cwt, accounting for 36.9% of the transactions. If the selling terms are FOB, the price range is \$0.58 lower than dressed delivery, with FOB at \$1.55/cwt compared to \$2.13/cwt for dressed delivery.

[5]In this study, the decision trees were created using the rpart package in R (Therneau et al., 2022).

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

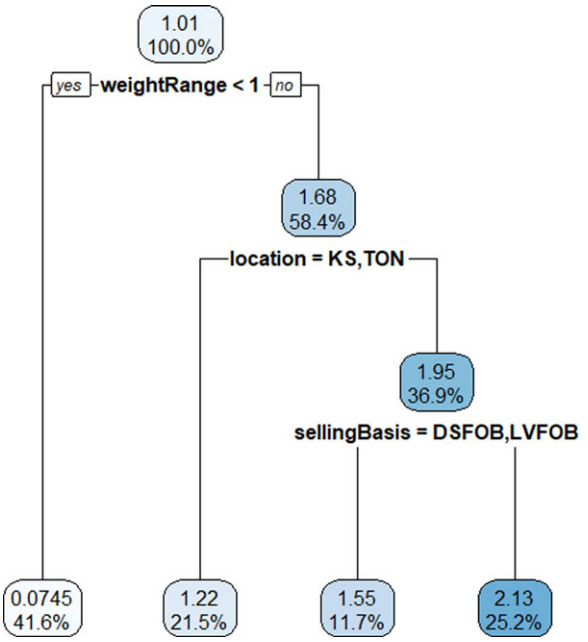

Figure 5. Decision tree results. Notes: Decision tree with default options, specifically cp = 0.01. At the root node, the average price range is $1.01/cwt. The first attribute used to split the dataset is the weightRange <1. When the weight range is less than 1 lb/head, 41.6% of the dataset is classified with the price range of $0.075/cwt. When the weight ranges are greater than 1 lb/head, the average price range is $1.68/cwt (about 58.4% of transactions), leading to further data partitioning based on the location and selling terms, where DS = dressed and LV = live.

have a price range of $1.95/cwt. In cases where the selling terms includes delivery and weight range exceeds 58 lbs/head, the price range further increases to $2.58, covering approximately 22.5% of transactions. When the weight range is below 58 lbs/head, the price range averages $1.76/cwt, accounting for about 11% of transactions.

Figure 6 illustrates the decision tree results employing the optimal cp value, leading to a more complex tree with additional variables for data stratification. This modification in the cp value results in more splits, providing a more detailed classification than the tree generated with the default cp value reported in Figure 5. Like Figure 5, the most significant factor influencing price range is the weight range, as indicated by the first split in the tree at weightRange <1. Among cattle with weightRange ≥1, location plays a crucial role, with significant differences observed between Kansas, Texas/Oklahoma/New Mexico and Nebraska. Cattle sold in Kansas, Texas/Oklahoma/New Mexico follow a different price range structure than those in Nebraska, highlighting regional market effects.

Further stratification reveals that for cattle sold in Kansas, Texas/Oklahoma/New Mexico, headcount and selling basis influence price range. Specifically, cattle lots with headCount <1,896 tend to have a lower price range. Additionally, if the sellingBasis is DSFOB or LVFOB, the price range increases further, particularly when weightRange <60. This trend suggests that different marketing and pricing mechanisms operate in different locations and selling conditions. Headcount further refines the segmentation, with cattle lots below 858 head in NE exhibiting greater price variation. When headcount is below 282, price range remains elevated, supporting the idea that smaller lot sizes tend to result in higher variability. Similarly, cattle with weightRange <60 in NE exhibit a significantly larger price range. These results highlight that weight and lot size are key determinants of price dispersion. In addition, as briefly mentioned above, the results in Figure 6 also find a location effect, with Kansas and Nebraska appearing on different branches of the tree. In case

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

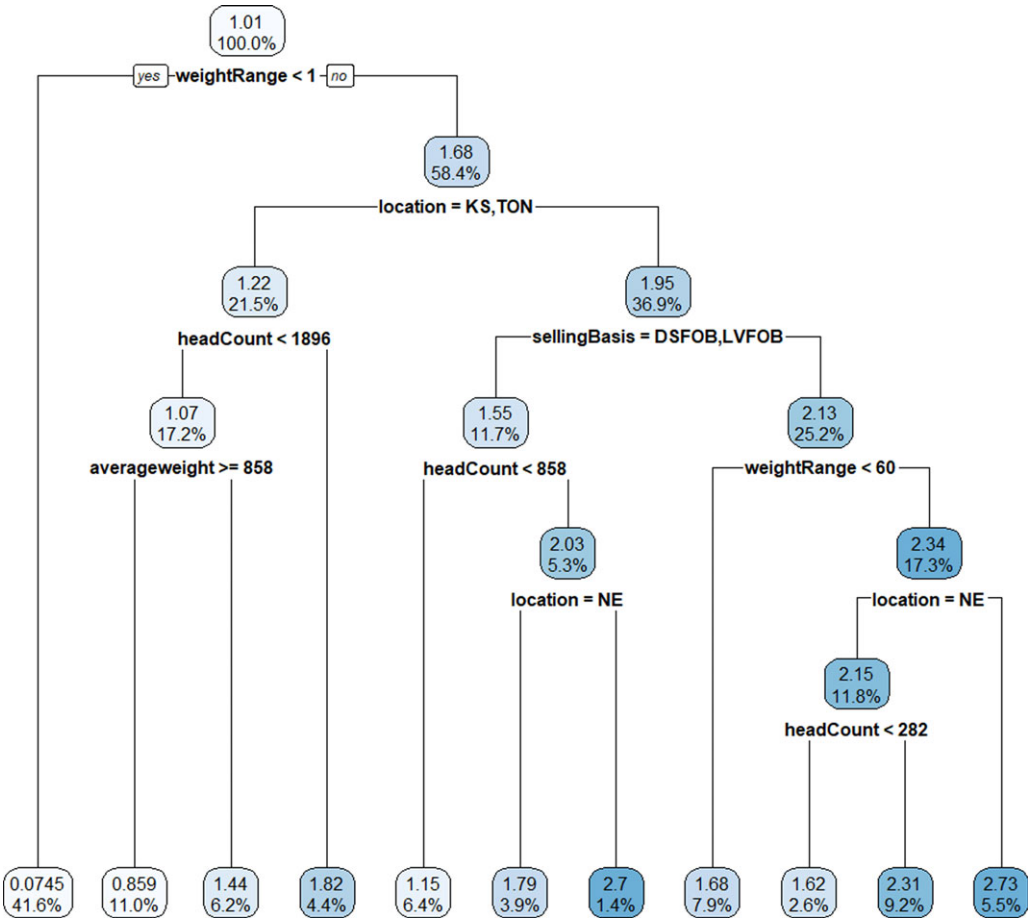

Figure 6. Decision tree results, cp = 0.003. Notes: The decision tree with the optimal complexity parameter (cp = 0.003) is presented. Similar to Figure 5, at the root node, the average price range is $1.01/cwt. The first attribute used to split the dataset is weightRange <1. When the weight range is less than 1 lb/head, 41.6% of the dataset is classified with a price range of $0.075/cwt. When the weight range exceeds 1 lb/head, the average price range increases to $1.68/cwt, representing approximately 58.4% of the transactions. This split leads to further partitioning based on attributes such as location, selling basis, head count, and others. Notably, when the weight range exceeds 60 lbs/head, the price range increases to $2.34/cwt, affecting 17.3% of the dataset.

of Kansas and Texas/Oklahoma/New Mexico, selling basis are not important factor influencing price range, which is not directly observable in the regression. However in other regions, Iowa/Minnesota and Nebraska, tends to have larger price range and also selling basis affects price range.

Overall, the results in Figures 5 and 6 indicate that an increase in the weight range, corresponding to more cattle, is associated with a larger price range. This pattern aligns with findings from Boyer et al. (2023), which reports the positive regression coefficients for head sold and weight range (Table 2 in Boyer et al. (2023)). Thus, a lower number of cattle sold within a day would likely result in less price variation. This suggests that heterogeneity within the cattle groups – such as variation in weight – plays an important role in price variation. Furthermore, the decision tree underscores that the FOB selling conditions generally lead to a larger price range compared to the dressed terms, which is consistent with the outcomes reported in Boyer et al. (2023). A notable distinction in the decision tree result is that weight range remains the primary factor influencing price range, with approximately 41.6% of transactions showing minimal price variation when the weight range is below 1 lb/head. This pattern is not explicitly observed in

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

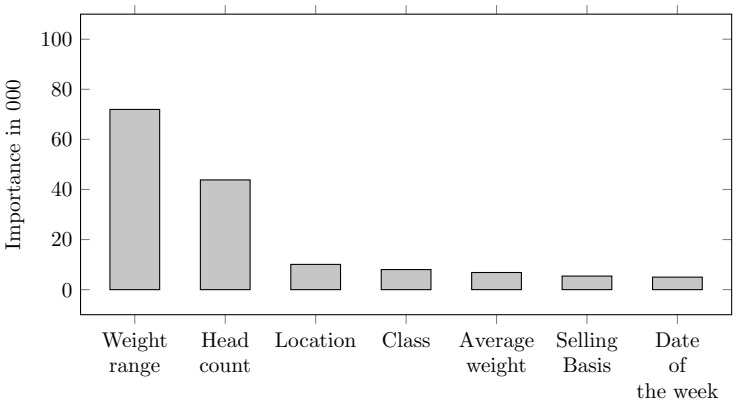

Figure 7. Important variables with decision tree in Figure 6. Note: Variable importance with cp = 0.003 is determined by calculating the relative influence of each variable: whether that variable was selected to split on during the tree building process, and how much the error in Equation (1) improved (decreased) as a result.

regression modeling but is evident in the decision tree analysis, emphasizing its importance in price stratification. In contrast, Boyer et al. (2023) identified selling terms as having a more pronounced effect on price variation.

In addition and divergent to Boyer et al. (2023), the decision trees analysis here did not reveal significant time-related influences; neither the day nor the month of trading emerged as significant factors in Figure 5 and Figure 6. Previous studies have noted an uneven distribution of cattle trade throughout the week, with lower activity on certain days. Boyer et al. (2023) found that the price range was lower on Tuesday, Wednesday, and Thursday compared to Friday, while Monday showed a higher price range than Friday. Marginal effects show that Monday has an approximate price range $0.07/cwt higher, while Tuesday, Wednesday, and Thursday have price ranges $0.40, $0.30, and $0.06 per cwt lower, respectively, compared to Friday. The lower price range on Tuesday might be due to reduced trading volume on that day (Boyer et al., 2023). Nonetheless, Figure 7 ranks the day of the week as the seventh most important variable.

5.2. Model validation and performance evaluation

The predictive performance of the decision tree models is assessed using several metrics: correlation, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). For the model depicted in Figure 5, the testing data results show a correlation of 0.530, MAE of 0.842, and RMSE of 1.365. These metrics indicate a moderate relationship between the predicted and actual price ranges, with the model showing a reasonable level of accuracy. In comparison, the model in Figure 6, which likely involves a different decision tree structure or tuning parameters, demonstrates slightly better performance metrics: a correlation of 0.568, MAE of 0.810, and RMSE of 1.324. This slight improvement in correlation and reduction in error metrics suggest that the model in Figure 6 captures the variability in the price range more effectively than the model in Figure 5.

To further assess the accuracy of these models, the predicted price ranges at various percentiles were compared to the actual price ranges, as shown in Figure 8. The analysis reveals that both models demonstrate robust performance at lower percentiles, accurately predicting price ranges between the 10th and 40th percentiles. However, notable discrepancies arise at higher percentiles, particularly from the 95th percentile onward. This divergence is attributed to the highly skewed distribution of price ranges, illustrated by the density plot where most data points are concentrated near zero, with a long tail extending toward higher values (Figure 4). This skewed distribution indicates that a significant proportion of observations have minimal price variation, which complicates the models' ability to predict extreme values accurately at the upper percentiles.

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

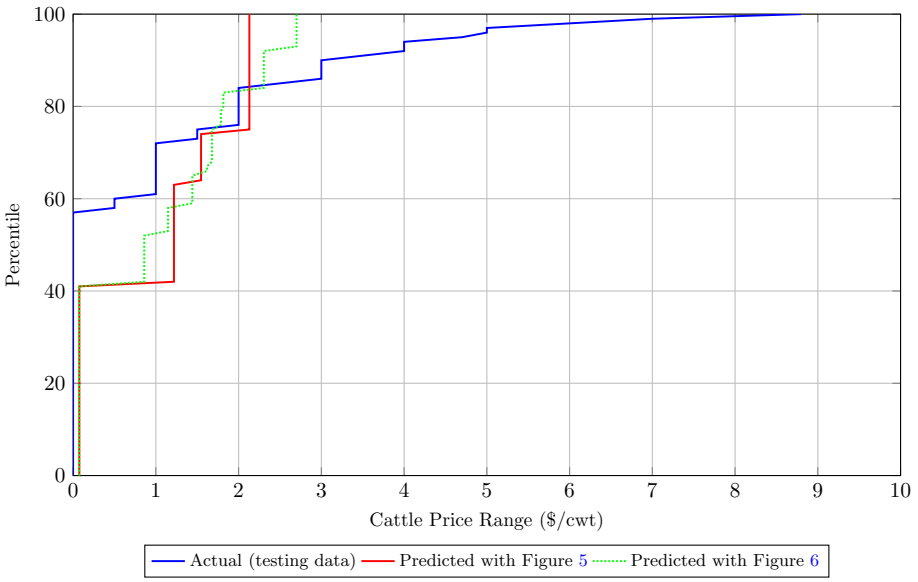

Figure 8. Cumulative distribution of actual and predicted cattle price ranges.

Overall, while both decision tree models offer reasonable predictions, the model depicted in Figure 6 shows slight improvements in both correlation and error metrics. Nevertheless, both models struggle with accurately predicting the most extreme price ranges, particularly above 95th percentile, indicating room for further refinement or the potential benefit of using more sophisticated models such as RF.

5.3. Random forest

In addressing the complexity of factors influencing cattle price ranges, traditional decision tree models, while insightful, often face limitations due to their susceptibility to overfitting and sensitivity to variations in the dataset. To overcome these limitations, we employ RF) model, introduced by Breiman (2001) and further developed by Liaw and Wiener (2002). RF enhances prediction accuracy and reduces overfitting by constructing an ensemble of decision trees, each trained on different subsets of the data. In the context of our study on cattle price ranges, RF's capability to handle numerous explanatory variables and diverse data types makes it particularly suitable. The RF model demonstrated improved performance, achieving a correlation of 0.635, MAE of 0.742, and RMSE of 1.243, indicating better predictive accuracy than the decision tree models. Additionally, the cumulative actual and predicted values are more closely aligned in the RF model, as illustrated in Figure 9 compared to the decision tree model shown in Figure 8. However, discrepancies still arise at higher percentiles, particularly from the 95th percentile onward.

In analyzing the factors influencing cattle price range, both decision tree and random forest models identify similar key variables but with nuanced differences in their significance. The decision tree model prioritizes weight range, head count, and average weight as the most critical factors, with weight range emerging as the dominant influence on price dispersion (Figure 7). These variables indicate that physical characteristics of cattle play a primary role in determining price variations. Other variables like selling basis, class, location, and day of the week also contribute but are less influential. The RF model, while also recognizing the importance of weight range, head count, and average weight, introduces additional factors such as month and grade as significant contributors (Figure 10). This suggests that seasonal trends and cattle quality grading have a substantial impact on price dispersion, factors that the decision tree model might have

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

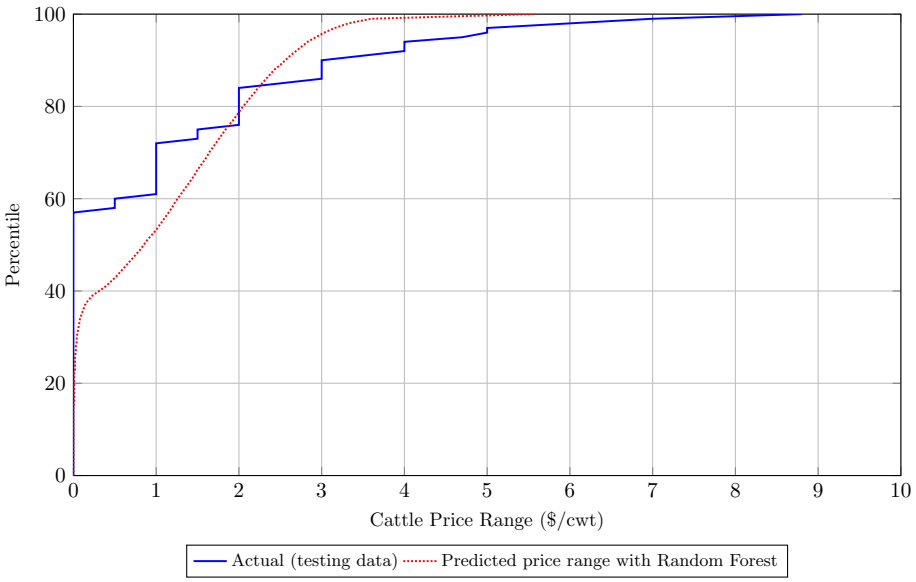

Figure 9. Cumulative distribution of actual and predicted cattle price ranges with Random Forest.

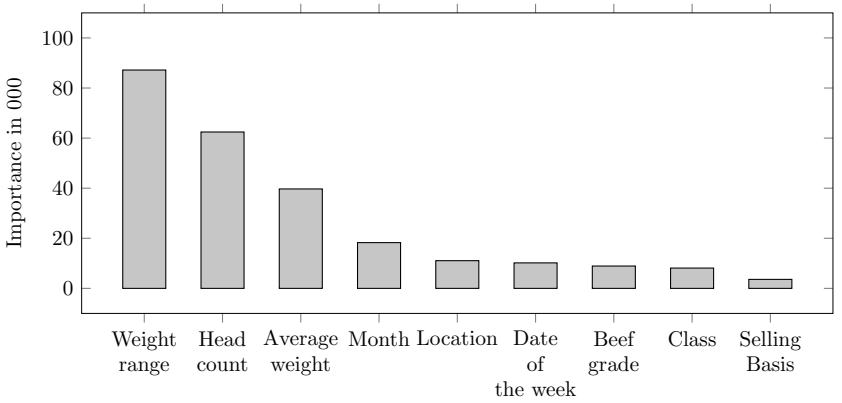

Figure 10. Important variables from Random Forest (RF). Note: Variable importance in the RF model, generated from 1,000 decision trees, is determined by evaluating the relative influence of each variable.

underrepresented. Additionally, location and day of the week remain relevant in the RF model, though their influence is less pronounced compared to the primary variables. Overall, while both models agree on the major drivers of price dispersion, the RF's enhanced ability to capture the complexity of the data highlights additional variables like month and grade, providing a more comprehensive understanding of the factors at play.

5.4. Removing zero price range

As previously discussed, approximately 41.6% of transactions exhibit a zero price range when the weight range is zero, which may affect our results. To address this, we removed all transactions where the price range is zero, leaving us with 57,452 observations for further analysis. We then re-applied the RF model and generated the importance of variables, as presented in Figure 11. After this adjustment, the weight range is still recognized as the most important factor, but we also observe that

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

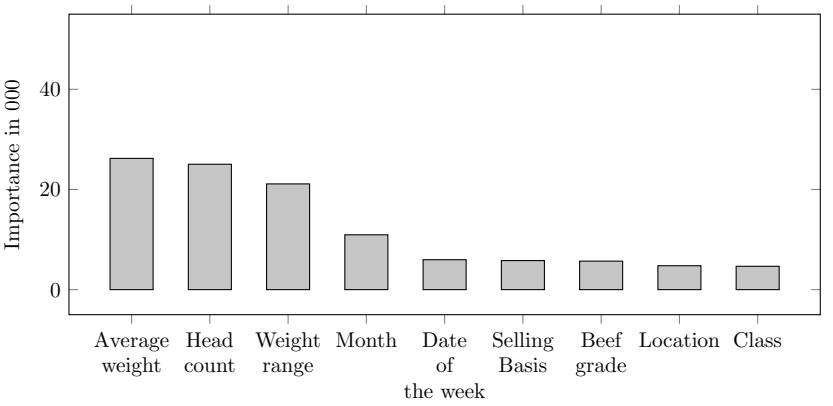

Figure 11. Important variables from Random Forest (RF) after removing zero price range observations. Note: Variable importance in the RF model, generated from 1,000 decision trees, is determined by evaluating the relative influence of each variable.

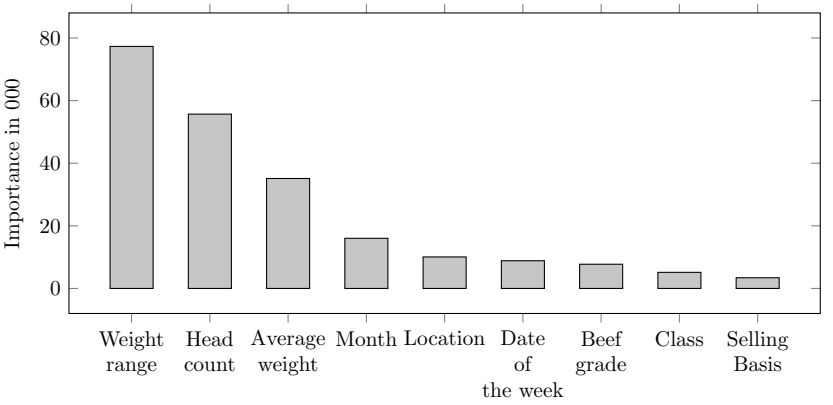

Figure 12. Important variables from Random Forest (RF) after excluding dairy cattle transactions. Note: Variable importance in the RF model, generated from 1,000 decision trees, is determined by evaluating the relative influence of each variable.

other factors, such as average weight and head count, become more influential in determining the price range (see Figure 11). Notably, even after removing all zero price range observations, the overall implications of the results remain consistent. This suggests that while weight range is important, other variables should also be considered when analyzing price range fluctuations.

5.5. Exclusion of dairy cattle transactions

Recognizing that dairy cattle and beef cattle prices may not be directly comparable due to significant differences in market dynamics – such as the size of cuts – we decided to exclude dairy cattle transactions to assess how this change would affect the results of our study. After excluding dairy cattle transactions, we re-applied the RF model solely to the fed cattle transactions data, consisting of 120,644 observations. The variable importance scores, calculated using 80% of the training data and displayed in Figure 12, revealed interesting shifts in the order of key factors influencing price variability. When dairy cattle transactions were included, the most important variable was weight range, followed by head count, average weight, month, and location. Other factors like day of the week (dow), grade, and class were still relevant, but their importance was lower in comparison.

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

However, after excluding dairy cattle transactions, the order of the important variables changed. Weight range remained the most influential variable, but head count and average weight gained greater prominence. That is while weight range continues to be the primary determinant of price variability, factors such as head count and average weight have a more significant role in the beef market once dairy transactions are excluded. The order of importance stayed at : weight range, head count, average weight, month, location, dow, grade, class, and selling basis.

6. Conclusion and policy implication

This study extends the findings of Boyer et al. (2023) by incorporating decision trees and RFs, machine learning techniques, as complementary tools to traditional regression analysis. These models excel in uncovering hidden information or patterns within data, providing additional insights into the variability of negotiated cash prices for fed cattle that go beyond what traditional regression models can reveal.

The study identifies key variables associated with fed cattle price ranges, such as weight range, head count, average weight, and transaction location. Notably, the weight range emerges as the primary variable influencing the price range, with smaller weight ranges linked to lower price ranges. Importantly, approximately 41.6% of transactions exhibit a zero price range when the weight range is zero, a nuance that was not evident in Boyer et al. (2023). The influence of transaction location, from the decision tree splits, is also noteworthy with transactions in Kansas and Nebraska generally displaying smaller price ranges compared to the other locations is also noteworthy. Additionally, RF analysis highlights the importance of other factors, such as month, grade, and selling basis, which were less emphasized in the decision tree model. For example, seasonal variations captured by the month variable and the impact of cattle quality grading on price dispersion offer a more comprehensive view of market dynamics. The decision tree approach, combined with RF insights, effectively captures the complex interplay between these variables, highlighting nuanced relationships that may not be evident through traditional regression modeling.

A significant contribution of this paper lies in finding that while weight range remains critical in determining price range, there are other relevant factors-ranked in order of relevancy-that also have an effect in price range. Moreover the analysis reveals that 41.6% of the dataset contains instances of zero weight range, making unclear the effect of this data in the estimated results. After removing transactions with a zero price range, weight range continues to be influential, and this effect remains significant and more so after also excluding dairy cattle from the dataset. These insights are particularly valuable for market participants and policymakers seeking to understand and manage price variability in the fed cattle market. These innovative and somewhat parsimonious methods are of considerable benefit for uncovering factors that impact price variability and found in this study. This approach demonstrates the potential for advanced analytical tools to mine data for explanatory power, offering stakeholders better and enhanced insights into the cattle market.

While this research provides valuable insights into the factors influencing fed cattle price dispersion, it is important to acknowledge and address its limitations. Decision trees and RFs, while effective in revealing variable importance and interactions, also present challenges in interpretation due to their complexity. Despite RF' advantage in reducing overfitting compared to decision trees, there remains a risk of overfitting. These limitations highlight the need for cautious interpretation of the results and suggest that further research is needed to refine the models and enhance their generalizability.

Data availability. The data that support the findings of this study are in the public domain and also available upon reasonable request.

Funding statement. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. Authors Wang, Kim, and Tejeda declare none.

https://doi.org/10.1017/aae.2025.10009 Published online by Cambridge University Press

CRediT. Wang, Zuyi: Conceptualization, formal analysis, Methodology, Software, Visualization, writing – original draft, writing – review and editing, Kim, Man-Keun: Conceptualization, formal analysis, Methodology, writing – original draft, writing – review and editing, Tejeda, Hernan: Conceptualization, Data Curation, formal analysis, Methodology, writing – original draft, writing – review and editing.

References

Adjemian, M., B. Brorsen, W. Hahn, T. Saitone, and R. Sexton. *Thinning Markets in US Agriculture. EIB-148*U.S. Department of Agriculture, Economic Research Service, 2016. https://doi.org/10.22004/ag.econ.232928, accessed on Jan 28, 2024.

Alpaydin, E. *Introduction to Machine Learning*. 3rd ed. Cambridge, Mass: The MIT Press, 2014.

Anderson, J., E. Clement, R. Stephen, S. Darrell, and N. James. "Experimental simulation of public information impacts on price discovery and marketing efficiency in the fed cattle markets." *Journal of Agricultural and Resource Economics* **23**(2019):262–79. https://www.jstor.org/stable/40986980

Azzam, A. "Market transparency and market structure: The livestock mandatory reporting act of 1999." *American Journal of Agricultural Economics* **85**,2(2003):387–95.

Boehmke, B., and B. Greenwell. *Hands-On Machine Learning with R*. Routledge, Taylor & Francis Group, 2020.

Boyer, C., C. Martinez, J. Maples, and J. Benavidez. "Price ranges from fed cattle negotiated cash sales." *Applied Economic Perspectives and Policy* **45**,3(2023):1–13 doi:10.1002/aepp.13296.

Breiman, L. "Random forest." *Machine Learning* **45**,1(2001):5–32 doi:10.1023/A:1010933404324.

Breiman, L., J. Firedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Wadsworth, 1983.

Chiu, Y.-W. *Machine Learning with R Cookbook*. Birmingham, UK: Packt Publishing Ltd, 2015.

Crespi, J.M., and R.J. Sexton. "Bidding for cattle in the Texas Panhandle." *American Journal of Agricultural Economics* **86**,3(2004):660–74.

Fan, C., M. Chen, X. Wang, J. Wang, and B. Huang. "Random forest." *Frontiers in Energy Research* **9**(2021):652801 doi:10.3389/fenrg.2021.652801.

Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer, 2009.

Lantz, B. *Machine Learning with R*. Birmingham, UK: Packt Publishing Ltd., 2013.

Liaw, A., and M. Wiener. "Classification and regression by random forest." *The R Journal* **2**,3(2002):18–22.

Pendell, D.L., and T.C. Schroeder. "Impact of mandatory price reporting on fed cattle market integration." *Journal of Agricultural and Resource Economics* **31**,3(2006):568–79.

Rokach, L., and O. Maimon. *Data Mining with Decision Trees*. Danvers, MA: World Scientific, 2014. https://doi.org/10.1142/9097.

Schroeder, J., R. Jones, J. Mintert, and A. Barkley. "The impact of forward contracting on fed cattle transaction prices." *Agribusiness: An International Journal* **15**,2(1993):325–337. https://www.jstor.org/stable/1349452.

Schroeder, T.C., G.T. Tonsor, and B.K. Coffey. "Commodity futures with thinly traded cash markets: the case of live cattle." *Journal of Commodity Markets* **15**(2019):1–15. https://doi.org/10.1016/j.jcomm.2018.09.005.

Schroeter, J.R., and A. Azzam. "Captive supplies and the spot market price of fed cattle: The plant-level relationship." *Agribusiness: An International Journal* **19**,4(2003):489–504.

Storm, H., C. Baylis, and T. Heckelei. "Machine learning in agricultural and applied economics." *European Review of Agricultural Economics* **47**,3(2020):849–92 doi:10.1093/erae/jbz033.

Therneau, T., B. Atkinson, and B. Ripley. (2022). An introduction to recursive partitioning using the RPART routines. Available at https://cran.r-project.org/web/packages/rpart, accessed on Jan 30, 2024.

Tibshirani, R. "Regression shrinkage and selection via the Lasso." *Journal of the Royal Statistical Society Series B* **58**,1(1996):267–88. https://www.jstor.org/stable/2346178.

USDA AMS. (2021). United States Department of Agriculture Agricultural Marketing Service. "3/70/20 Confidentiality Guidelines. Available at https://www.ams.usda.gov/sites/default/files/LMRConfidentialityGuidelinePresentation.pdf, accessed on Feb 24, 2024.

Ward, C.E. "Inter-firm differences in fed cattle prices in the Southern Plains." *American Journal of Agricultural Economics* **74**,2(1992):480–5.

Ward, C.E., S.R. Koontz, and T.C. Schroeder. "Impacts from captive supplies on fed cattle transaction prices." *Journal of Agricultural and Resource Economics* **23**,2(1998):494–514.

Ziegler, A., and I. König. "Mining data with random forests: current options for real-world applications." *WIREs Data Mining and Knowledge Discovery* **4**,1(2014):55–63 doi:10.1002/widm.1114.

Dr. Zuyi Wang is a postdoctoral research fellow in the Department of Agricultural Sciences at Clemson University.

Dr. Man-Keun Kim is a Professor and Director of Graduate Studies in the Department of Applied Economics at Utah State University.

Dr. Hernan Tejeda is an Associate Professor and Extension Specialist at University of Idaho.

Cite this article: Wang, Z., M.-K. Kim, and H. Tejeda (2025). "Unraveling Hidden Patterns in Fed Cattle Negotiated Cash Prices Using Machine Learning." *Journal of Agricultural and Applied Economics*. https://doi.org/10.1017/aae.2025.10009