Journal of Agricultural and Applied Economics (2026), 1–19
doi:10.1017/aae.2026.10050

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Predicting Thoroughbred Yearling Auction Prices with Machine Learning: Evidence from the Keeneland September Sale

Yanchao Yang[1], John Clarke[2], Tapan Mandal[1], Thinh Nguyen[1] and Trung Pham[1]

[1]School of Business and Leadership, DePauw University, USA and [2]EMR Metal Recycling, USA
Corresponding author: Yanchao Yang; Email: yyang056@ucr.edu

Abstract
We apply machine learning methods to predict Thoroughbred yearling auction prices at the Keeneland September Sale (2020–2024). Our sample includes 5,788 yearling prices with pedigree data. We use both linear and tree-based models to predict log prices. We use cross-validation to tune model hyperparameters and select Ridge regression ($\alpha = 1.451$) as the primary model for interpretation given its stability and interpretability. The Ridge regression explains approximately 54% of out-of-sample variation ($R^2 \approx 0.5403$). Sire and Dam Reputation emerge as the dominant predictors. Results provide pricing benchmarks and show how reputation and session structure shape Thoroughbred yearling auction prices.

Keywords: auction price prediction; bloodstock markets; machine learning; regularized regression; thoroughbred yearlings
JEL classifications: C45; C55; G12; Q19; L83

1. Introduction

Livestock and animal-asset auctions provide a valuable empirical setting for studying price formation under uncertainty, information asymmetry, and heterogeneous quality. Prior studies on livestock and animal auctions show that animal traits, pedigree signals, auction management, and seller reputation are all associated with auction prices (Calil et al., 2022; Maynard and Stoeppel, 2007; Rogers et al., 2023; Schroeder et al., 1988; Schulz et al., 2015; Zimmerman et al., 2012). Within this animal-asset pricing market, Thoroughbred yearling auctions represent a market with substantial transaction volume. Major sales such as the Keeneland September Yearling Sale generated over $405 million in 2024 (Keeneland Association, 2024). More broadly, the US equine industry contributes an estimated $177 billion in total economic impact and involves over 6.5 million horses nationwide (American Horse Council, 2023).

Unlike older racehorses, yearlings are sold before they have any racing record, so buyers face substantial uncertainty and asymmetric information in the auction market (Chezum and Wimmer, 1997). Yearling prices largely reflect expectations about future potential rather than demonstrated performance. Prior studies have examined this pricing mechanism using hedonic price models. Studies have shown that sire stud fees, dam performance, and session placement are all associated with price variation in Thoroughbred yearling auctions (Buzby and Jessup, 1994; Chezum and Wimmer, 1997; Plant and Stowe, 2013). However, to the best of our knowledge, the use of modern data science techniques such as supervised machine learning remains rare in this area. The high dimensionality of pedigree signals and the potential for nonlinearity among reputational signals make this setting well suited for machine learning approaches (Athey and

Imbens, 2019; Varian, 2014). Moreover, machine learning framework emphasizes out-of-sample prediction, which can provide a more credible prediction for new auction prices (Hastie et al., 2009; Mullainathan and Spiess, 2017). To date, few studies have systematically applied machine learning methods to yearling price prediction.

This study uses auction data from the Keeneland September Yearling Sale, combined with pedigree information, to predict the prices of Thoroughbred yearlings using a machine learning framework. Using data from 2020 to 2024, our main analytical sample includes 5,788 yearling auction prices and pedigree information such as sire and broodmare sire's progeny records, sire–broodmare sire cross performance, dosage profiles, and sire and dam reputation metrics. We use a set of linear models (OLS, Ridge, Lasso, Elastic Net, and Huber) and tree-based models (Random Forest, XGBoost, LightGBM, and CatBoost), along with a weighted ensemble of the top-performing tree models, to predict log auction prices. Among these models, we use Ridge regression for the session-level analysis because it offers greater stability and interpretability.

Our analysis reveals three key findings. First, the Ridge model ($\alpha \approx 1.451$) explains approximately 54% of the out-of-sample variation in auction prices ($R^2 \approx 0.5403$). Second, reputation signals are among the strongest predictors, with Sire Reputation emerging as the dominant predictor. Dam Reputation is also economically important, even though its estimated coefficient is smaller. This difference likely reflects measurement precision rather than true economic importance. Sires produce many more progeny, so their market reputation can be estimated more precisely. Dams have far fewer progeny, making the coefficients of Dam Reputation noisier. Third, model performance varies across session segments, with earlier sessions (Sessions 2 and 3) showing session-level R^2 up to 0.3725, while later sessions (Sessions 7 and 11) exhibit weaker fit, with R^2 near 0.21.

Several limitations should be noted. First, the Sire and Dam Reputation indices are constructed from historical auction outcomes, raising concerns about circularity and endogeneity. Coefficient estimates should be interpreted as predictive associations rather than causal effects. Second, our analysis uses data from a single auction and might have limited external validity. Our results may not fully generalize to other yearling auctions such as Fasig-Tipton or Ocala Breeders' Sales Company (OBS). Third, some session-level analyses are based on small sample sizes, which may reduce statistical precision. Overall, our findings are best viewed as documenting robust predictions for yearling auction prices rather than establishing causal mechanisms.

The remainder of the paper is organized as follows. Section 2 provides background on auction theory and prior research. Section 3 details our data and methodology. Section 4 presents model results and interpretation. Section 5 discusses the implications of our findings, and Section 6 concludes.

2. Background

A substantial literature in livestock and animal auctions examines how observable characteristics, such as management practices, genetic merit, and reputation signals are associated with prices. In cattle markets, hedonic models show that physical traits, health protocols, seller reputation, and market conditions significantly influence auction prices (G. F. Schroeder et al., 1988; Schulz et al., 2015). More recent work highlights the role of genetic indicators and value-added attributes in seedstock and breeding markets, showing that pedigree signals, expected progeny differences, and farm reputation are reflected in prices under conditions of uncertainty (Calil et al., 2022). Research on market transparency further emphasizes how information structure shapes price discovery and the extent to which observed attributes explain transaction-level variation (Rogers et al., 2023; Schroeder et al., 2023).

Predicting Thoroughbred yearling prices is uniquely challenging. Because these one-year-old horses have no race performance history, buyers make auction decisions under asymmetric

information, relying mainly on pedigree, physical conformation, sale timing, and cataloged information to assess future potential (Chezum and Wimmer, 1997). Theoretical frameworks from auction economics further suggest that buyers' valuations are shaped not only by private signals but also by expectations about competitor behavior, creating room for winner's curse and adverse selection (Chezum and Wimmer, 1997; Maynard and Stoeppel, 2007; Milgrom and Weber, 1982). Prior studies also explore disclosure effects within yearling auctions, suggesting that voluntary catalog disclosure and strategic withholding of information may meaningfully affect price signals (Plant and Stowe, 2013).

Other studies identify factors such as sire stud fee, dam racing record, foaling date, sex, and catalog placement as important drivers of auction price variation (Buzby and Jessup, 1994; Chezum and Wimmer, 1997; Plant and Stowe, 2013). Maynard and Stoeppel (2007) examine pricing for Thoroughbred broodmares and find that the performance of the mare's existing foals is a stronger price determinant than her own racing record. Vickner and Koch (2001) examine yearling prices using a hedonic model to estimate the marginal value of pedigree and auction characteristics. Their results show that sire quality, stud fees, and dam-related traits are significant determinants of price. Parsons and Smith (2007) extend this approach to the British yearling market by examining how catalog placement and eligibility for owners' premium programs affect auction prices. These premium programs provide additional payments to owners of qualifying horses, thereby increasing the expected value of eligible yearlings. More recently, Mouncey et al. (2024) analyze UK yearling sales data and find that sire covering fee, dam performance, consignment size, and catalog session all play significant roles in pricing. Kim et al. (2019) apply hedonic models to Australian yearling auctions and highlight the predictive value of stud fees, sibling race performance, and sales history in explaining auction prices. Kibler and Thompson (2020) examine stock-type horses sold online and find that digital presentation strategies, particularly videos, boost sale prices by around 11%, reflecting changing buyer behavior in tech-mediated auctions.

Prior bloodstock auction studies provide useful insights, but they have several limitations. First, most rely on linear regression models, which are interpretable but may be too restrictive for modeling yearling auction prices. Second, much of the prior work focuses on in-sample explanation rather than out-of-sample prediction, which lacks assessments on how well the models perform on new auction data. Third, many studies use older data, while recent changes in auction dynamics, richer catalog information, and advances in predictive modeling call for updated research.

Recent studies have applied machine learning methods to other livestock auctions. Haque et al. (2021) use supervised learning methods to predict prices of cows, finding that ensemble models significantly outperform linear regressions. Zapata and Anderson (2025) introduce an approach based on Data Envelopment Analysis (DEA) to quantify prices in cattle auctions and recommend cost-effective improvements to animal traits. Wang et al. (2025) apply decision trees and random forests to fed-cattle cash prices and reveal weight range as a primary determinant of price variability. Prior studies show that machine learning methods can complement traditional models by capturing nonlinear pricing patterns and improving price prediction in livestock auctions. This suggests that similar approaches may also be useful for predicting Thoroughbred yearling prices.

3. Methodology

3.1. Data collection and processing

The dataset used in this analysis combines publicly available sale records from the Keeneland September Yearling Sale (2020–2024) with pedigree data obtained from Equineline.com, a service operated by The Jockey Club Information Systems (TJCIS). Keeneland sales provide transaction-level data for each horse offered at auction, including sale price, hip number, sex, color, session

placement, and consignor information. These data are collected directly from Keeneland's official website archives. Pedigree data from Equineline's Free 5-Cross Pedigree reports include lineage information for the sire, dam, and broodmare sire, along with relevant performance metrics including number of foals, starters, winners, black-type winners[1], and total progeny earnings. Horses are matched across datasets using registered name and year of sale.

Data preprocessing involves several standardization steps. Column names are cleaned by removing extra whitespace, currency strings are converted to numeric values, and invalid entries are converted to missing values and imputed using robust statistics. Seller, sire, and dam names are standardized by converting them to lowercase and removing organizational suffixes, agent qualifiers, and country codes to ensure consistent entity identification across sources. Following initial preprocessing, we filter high-influence observations using Cook's distance as a diagnostic measure for outlier detection (Cook, 1977). For predictors with highly skewed distributions, we apply monotone power transformations from the Yeo–Johnson or Box–Cox families (Box and Cox, 1964; Yeo and Johnson, 2000). These transformations are used when they help make the data more symmetric while keeping the results easy to interpret. Using these approaches, we identify and remove 272 outliers, resulting in a final analytical sample of 5,788 observations.

3.2. Outcome variable and predictors

The outcome variable in this study is the natural logarithm of the yearling's final sale price. This transformation addresses the substantial right skewness in raw prices and helps stabilize variance across the price range. Our predictors fall into four broad categories: reputation measures, progeny-performance measures, dosage-profile variables, and additional control variables.

We first construct Sire and Dam Reputation indices from four measures of offspring sale performance, following Alani (2006). The first measure is the 90th percentile of a sire's or dam's offspring log sale prices (weight 0.40), which captures the high end of their sale performance. The second is the share of offspring selling above the market-wide 90th-percentile price (weight 0.30), capturing the frequency of top-tier sales. The third is a sire's or dam's median offspring log sale price, expressed as a percentile rank among all sires or dams (weight 0.20), representing typical market standing. The fourth is the 25th percentile of offspring log sale prices (weight 0.10), which reflects the lower end of the sale distribution. The declining weights place greater emphasis on elite sale outcomes and still account for consistency across the full distribution of offspring sales. Each measure is min–max scaled prior to aggregation, producing an index ranging from zero to one, where higher values indicate stronger market reputation.

We include several progeny-performance variables to measure pedigree quality. For the sire, we use the sire's total number of foals and the sire's black-type winner rate, defined as the share of the sire's offspring that became elite winners. For the broodmare sire, we use the broodmare sire's total number of foals and the broodmare sire's black-type winner rate, defined in the same way. We also measure the performance of the specific sire–broodmare sire cross using Average Earning Index (AEI) and total earnings. AEI captures the earnings performance of offspring from that specific cross relative to the typical racehorse, while total earnings measure the cumulative earnings of offspring from that cross.

The dosage profile includes five indices (Brilliant, Intermediate, Classic, Solid, and Professional) that summarize the pedigree's balance between speed and stamina. Additional control variables include year fixed effects (2021–2024, with 2020 as the reference year), sex, and first-crop status, which indicates whether the yearling is from a sire's first crop of offspring.

We evaluate each candidate predictor using four criteria and combine the resulting scores to select variables most strongly associated with price. The F-test captures the strength of each

[1]Black-type winner indicates a horse has placed in the top three of a stakes race, with its name highlighted in bold, black font in sales catalogs.

variable's linear relationship with price, while mutual information measures the reduction in uncertainty about price associated with each variable. Lasso regression estimates the model using all candidate predictors and shrinks weaker coefficients toward zero. Elastic Net regression combines shrinkage with grouped selection, allowing correlated predictors to be retained together. Predictors that perform well across these four criteria are included in the final predictor set. This selection process reduces our initial 29 variables to a final set of 20 features.

3.3. Model selection and estimation strategy

We estimate two classes of models – linear models and tree-based models – to evaluate predictive performance, stability, and interpretability. We use linear models to provide transparent coefficient estimates and use tree-based models to capture potential nonlinear relationships and higher-order interactions.

We estimate five linear models. Ordinary least squares (OLS) serves as the baseline model without regularization. Ridge regression applies an L2 penalty to shrink coefficients toward zero. Lasso regression uses an L1 penalty to both shrink coefficients and perform automatic variable selection. Elastic Net combines L1 and L2 penalties to balance sparsity and shrinkage. Huber regression provides a robust alternative by down-weighting the influence of outliers, thereby reducing sensitivity to extreme observations. For models requiring hyperparameter tuning, we use cross-validation to identify optimal values. We divide the training data into five folds and repeat this process three times with different random splits, creating 15 total validation runs. For Ridge and Lasso models, we test penalty parameters (α) ranging from 10^{-5} to 10^{3} on a logarithmic scale. For Elastic Net, we also test mixing parameters from 0.10 to 0.99 to determine the balance between L1 and L2 penalties. Before model estimation, we standardize all continuous variables using robust scaling followed by standard scaling to ensure comparable coefficient magnitudes.

For tree-based models, we estimate Random Forest, XGBoost, LightGBM, and CatBoost and additionally construct a weighted ensemble. We tune both structural and regularization parameters to control model complexity and reduce overfitting. For Random Forest, we tune the maximum tree depth, the minimum number of observations required in a terminal leaf, the minimum number of observations required to split an internal node, and the number of predictors randomly considered at each split. These parameters are searched over bounded grids to limit excessive tree growth while preserving the model's ability to capture nonlinear relationships and interactions. For XGBoost, which includes explicit regularization, we tune both L1 and L2 penalties on logarithmic scales. The L1 penalty is searched from 10^{-8} up to 10^{2} on a logarithmic scale, and the L2 penalty from about 10^{-3} up to 10^{2} on a logarithmic scale. In addition, we tune structural complexity parameters such as the minimum loss reduction required to make a split, maximum tree depth, and the minimum child weight. For LightGBM, we similarly tune L1 regularization over a log scale from roughly 10^{-8} to 10^{2} on a logarithmic scale and L2 regularization from approximately 10^{-3} to 10^{2} on a logarithmic scale. We also tune the number of leaves, the minimum number of observations required in each leaf, and subsampling parameters controlling the proportion of features and observations used in each iteration. For CatBoost, we tune the L2 leaf regularization parameter over a range from about 1 to 50 depending on the grid, along with tree depth, learning rate, and iteration settings.

3.4. Model evaluation

We split the dataset into three parts: 60% for training, 20% for validation, and 20% for prediction evaluation. To maintain representativeness across all splits, we stratify the data by session, ensuring that each split contains similar proportions of yearlings from different sale sessions. We evaluate model performance using four metrics. R^2 measures the proportion of variance in log

prices explained by the model. Adjusted R^2 modifies regular R^2 by accounting for the number of predictors relative to sample size, providing a more conservative estimate of model fit when comparing models with different numbers of variables. Root mean squared error (RMSE) measures the magnitude of prediction errors in outcome variables, giving greater weight to larger errors. Mean absolute error (MAE) represents the average absolute prediction error, treating all deviations equally and making it less sensitive to outliers than RMSE. For each metric, we also report the standard error to assess model stability.

For coefficient estimates of linear models, we use bootstrap resampling as the primary inferential framework. We draw 1,000 samples with replacement from the training data, each maintaining the original sample size. For each sample, we re-estimate the model and record the coefficients. This process generates distributions for each coefficient, from which we calculate 95% confidence intervals and determine statistical significance.

Among all models, we select Ridge regression for coefficient interpretation and session-level analysis for its predictive performance stability and coefficient interpretability. To verify that the Ridge specification meets statistical assumptions, we conduct several diagnostic tests. We test residual normality using the Jarque–Bera, Shapiro–Wilk, and Kolmogorov–Smirnov tests (Jarque, 1980; Shapiro, 1965). We check for autocorrelation using the Durbin–Watson statistic (Durbin and Watson, 1950; Godfrey, 1978). We assess constant variance assumptions using the Breusch–Pagan and White tests (Breusch and Pagan, 1979; White, 1980). We present the cross-validated α tuning curve and corresponding coefficient shrinkage (regularization) path for the Ridge specification in the Appendix (Figure A2 and Figure A3). We also show the correlation matrix in Figure A1.

To examine heterogeneity across auction sessions, we evaluate model performance separately by session. For each session, predictive accuracy is summarized using R^2, and sessions are ranked by R^2 to assess whether model performance differs across market segments. We also estimate separate Ridge regressions by session to examine whether the associations between key predictors and yearling prices vary across different sessions of the auction.

In addition, we conduct several robustness checks. Because the 2020 sale occurred during the COVID-19 pandemic, we first test whether the 2020 price distribution differs materially from those of subsequent years using a Kruskal–Wallis test with pairwise comparisons. We then evaluate temporal prediction stability through a leave-one-year-out cross-validation (LOYO-CV) procedure, in which the model is iteratively trained on four years of data and evaluated on the held-out year. Finally, we re-estimate the Ridge model after excluding all 2020 observations and assess its predictive performance separately. Together, these procedures allow us to determine whether the main findings are sensitive to the inclusion of the pandemic-year observations or to temporal instability across sale years.

4. Results

4.1. Descriptive analysis

Table 1 presents descriptive statistics for the outcome variables and predictors in this study. Sale prices exhibit substantial right skewness characteristic of auction markets. The mean price of $128,428 exceeds the median of $60,000, with individual transactions ranging from $1,000 to $2.5 million. After transformation, log prices show improved symmetry with a mean of 10.94 and standard deviation 1.37.

The average number of offspring per sire is 856 (SD = 662), indicating substantial variation, with some sires producing many offspring and others relatively few. The Sire's Black-type Winner rate is 6%. The average number of offspring per broodmare sire is 1,930 (SD = 1,640). The Broodmare Sire's Black-type Winner rate is 5%. The AEI for sire-broodmare sire cross has a mean of 1.17, suggesting that, on average, the combinations in this sample produce offspring

Table 1. Summary statistics for auction prices, pedigree performance, dosage profile, and reputation metrics

Variable	Definition	Mean	Std. dev.	Median	Min.	Max.
Outcome variables						
Price	Sale price (USD)	128,428	185,712	60,000	1,000	2,500,000
Log price	Log-transformed price	10.94	1.37	11	6.91	14.73
Sire's progeny record						
Sire foals	Sire's total number of offspring	856	662	680	5	3,921
Sire BW rate	Sire's rate of offspring who become elite winners	0.06	0.03	0.06	0	0.14
Broodmare sire's progeny record						
BS foals	Broodmare sire's total number of offspring.	1,930	1,640	1,500	1	9,167
BS BW rate	Broodmare sire's rate of offspring who become elite winners	0.05	0.02	0.05	0	0.25
Sire × Broodmare sire cross performance						
AEI	Earnings index for the specific sire/broodmare sire cross	1.17	1.35	0.89	0	19.34
Total earnings	Total earnings for the specific sire/broodmare sire cross	656,496	1,567,583	146,325	0	30,970,919
Dosage profile						
Brilliant	Speed influence	2.93	1.9	3	0	15
Intermediate	Speed-stamina blend	4.68	3.38	4	0	22
Classic	Stamina influence	6.64	3.72	6	0	22
Solid	Extended stamina	0.43	0.97	0	0	12
Professional	Extreme stamina	0.02	0.17	0	0	3
Reputation metrics						
Sire reputation	Sire's commercial prestige index	0.51	0.18	0.49	0	0.98
Dam reputation	Dam's commercial prestige index	0.35	0.11	0.31	0.12	0.97

earning 17% more than the typical racehorse. However, the high standard deviation (SD = 1.35) and a median value of 0.89 suggest that the mean is likely elevated by a small number of exceptionally successful sire–broodmare sire pairings. This skewness is also evident in the total earnings for these crosses, which average $656,496 but have a median of only $146,325.

The dosage profile for yearlings serves as a pedigree signal, indicating a market preference for horses with a balanced combination of speed and stamina traits. The Classic category shows the highest average scores. In contrast, the average of Professional, which shows the extreme stamina, is the smallest. The reputation metrics, which serve as proxies for commercial prestige, show that Sire Reputation averages 0.51 on a 0-1 scale while Dam Reputation averages 0.35.

Figure 1 plots the distribution of sale prices across the 12 sale sessions and reveals a clear pattern of market stratification. A downward trend is evident in both the median price and overall price dispersion as the sale progresses from the earliest to the later sessions. Sessions 1 and 2, which represent the most selected segments of the auction, exhibit the highest median prices – approximately $400,000 or more – and the widest interquartile ranges. These early sessions also contain many high-value outliers, with several yearlings selling for over

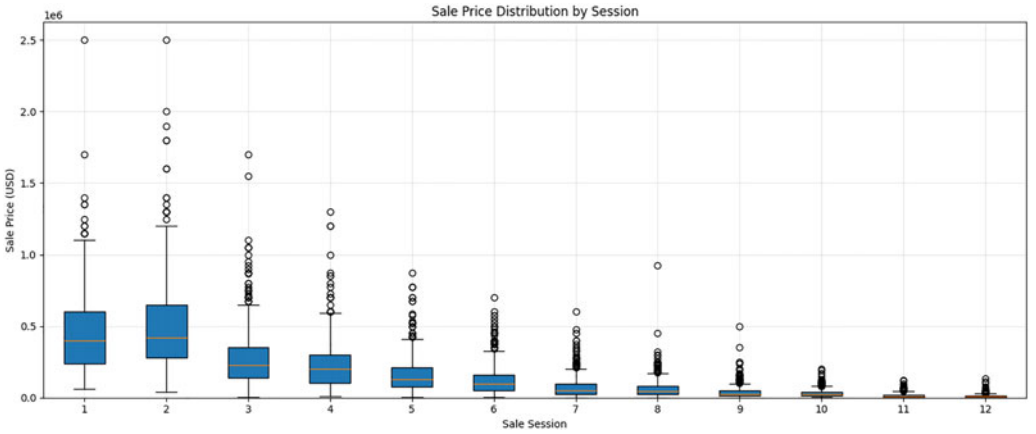

Figure 1. Sale price distribution by session.

$1.5 million. Conversely, beginning around Session 5, there is a decline in both price levels and dispersion. The median price in the final sessions falls below $50,000, and the price distribution becomes much more compressed. This tiered market structure highlights the role of sale session as a key determinant of price and underscores the importance of examining session-level predictive performance and the relationships between predictors and auction prices.

Appendix Table A3 reports the number of unique sires and dams by year and session. Our analytical sample includes 279 unique sires and 4,807 unique dams. Each dam contributes exactly one yearling per year, so the number of dams equals the annual sample size. In contrast, a single sire can produce multiple yearlings each year, which leads to fewer unique sires across years. These patterns reflect that sires have many more progenies than individual dams. Each sire's reputation is estimated from multiple progenies, which reduces sampling variability and increases precision. In contrast, each dam typically contributes only one progeny per year, limiting the variation used to estimate Dam Reputation and making these estimates less precise relative to Sire Reputation.

4.2. Machine learning results

Table 2 shows that linear models – OLS, Ridge, Lasso, Elastic Net, and Huber – deliver very similar predictive performance, with test R^2 values ranging between 0.5400 and 0.5403. Tree-based and ensemble methods perform somewhat differently: the Weighted Ensemble and Random Forest achieve the highest predictive accuracy, with test R^2 values of 0.5503 and 0.5533, respectively, while CatBoost, XGBoost, and LightGBM perform less strongly in this setting. These results indicate that the predictive gains from more complex nonlinear models are modest. Given this pattern, model selection requires consideration of other criteria beyond predictive accuracy. Therefore, we select Ridge regression ($\alpha = 1.451$) based on three criteria. First, Ridge effectively shrinks correlated coefficients while retaining all predictors, which is critical for preserving economically meaningful variables that Lasso might eliminate when handling multicollinearity (Dormann et al., 2013). Second, Ridge regression provides transparent coefficient estimates that are better suited for economic interpretation than tree-based or ensemble methods. Third, Ridge has smaller standard errors for R^2, Adjusted R^2, RMSE, and MAE, indicating more stable out-of-sample performance. Therefore, Ridge is adopted as the baseline model for inference, interpretation and session-level analysis. Other models are best viewed as robustness checks on predictive performance.

Table 3 provides the technical specifications for Ridge model. The regularization parameter $\alpha = 1.451$ represents the optimal penalty strength identified through cross-validation, balancing prediction accuracy against coefficient stability. This moderate regularization level indicates

Table 2. Model performance comparison

Model	R^2 (SE)[a]	Adjusted R^2 (SE)[b]	RMSE (SE)[c]	MAE (SE)[d]
Ordinary least squares	0.5400 (0.0075)	0.5264 (0.0077)	0.9289 (0.0065)	0.7428 (0.0058)
Ridge	0.5403 (0.0075)	0.5266 (0.0077)	0.9287 (0.0065)	0.7425 (0.0058)
Lasso	0.5402 (0.0075)	0.5266 (0.0077)	0.9287 (0.0065)	0.7426 (0.0058)
Elastic net	0.5403 (0.0075)	0.5266 (0.0077)	0.9287 (0.0065)	0.7425 (0.0058)
Huber	0.5401 (0.0075)	0.5264 (0.0077)	0.9289 (0.0066)	0.7422 (0.0058)
Weighted ensemble	0.5503 (0.0091)	0.5369 (0.0094)	0.9182 (0.0077)	0.7245 (0.0070)
CatBoost	0.5322 (0.0096)	0.5183 (0.0099)	0.9364 (0.0078)	0.7414 (0.0064)
LightGBM	0.4934 (0.0101)	0.4784 (0.0104)	0.9745 (0.0082)	0.7691 (0.0069)
Random forest	0.5533 (0.0085)	0.5400 (0.0088)	0.9153 (0.0079)	0.7248 (0.0071)
XGBoost	0.5153 (0.0098)	0.5009 (0.0101)	0.9533 (0.0085)	0.7530 (0.0066)

[a]Out-of-sample R^2across cross-validation folds; the standard error reflects variability across folds.
[b]Adjusted R^2accounting for the number of predictors; the standard error reflects variability across folds.
[c]Root mean squared error averaged across cross-validation folds; measures the typical magnitude of prediction errors in the outcome's units. The standard error reflects variability across folds.
[d]Mean absolute error averaged across cross-validation folds; captures the average absolute prediction error. The standard error reflects variability across folds.

Table 3. Ridge model specification

Parameter	Description	Value
α (alpha)	L2 regularization parameter	1.451
Predictors	Number of features retained	20
Training N	Observations for model fitting	3,472
Cross-validation R^2	15-fold CV performance	0.5403 (±0.0279)

sufficient shrinkage to handle multicollinearity without over-penalizing coefficients. The training sample of 3,472 observations (60% of the analytical sample) provides adequate statistical power for reliable coefficient estimation. The cross-validation R^2 of 0.5403 (± 0.0279) demonstrates consistent performance across different data folds.

4.3. Estimated coefficients and market implications

Figure 2 reports on the Ridge regression coefficients and their statistical significance levels. Sire Reputation dominates all other predictors, with a coefficient of 1.001 (significant at the 1% level). This dominance of paternal lineage confirms findings by Hansen and Stowe (2018) that

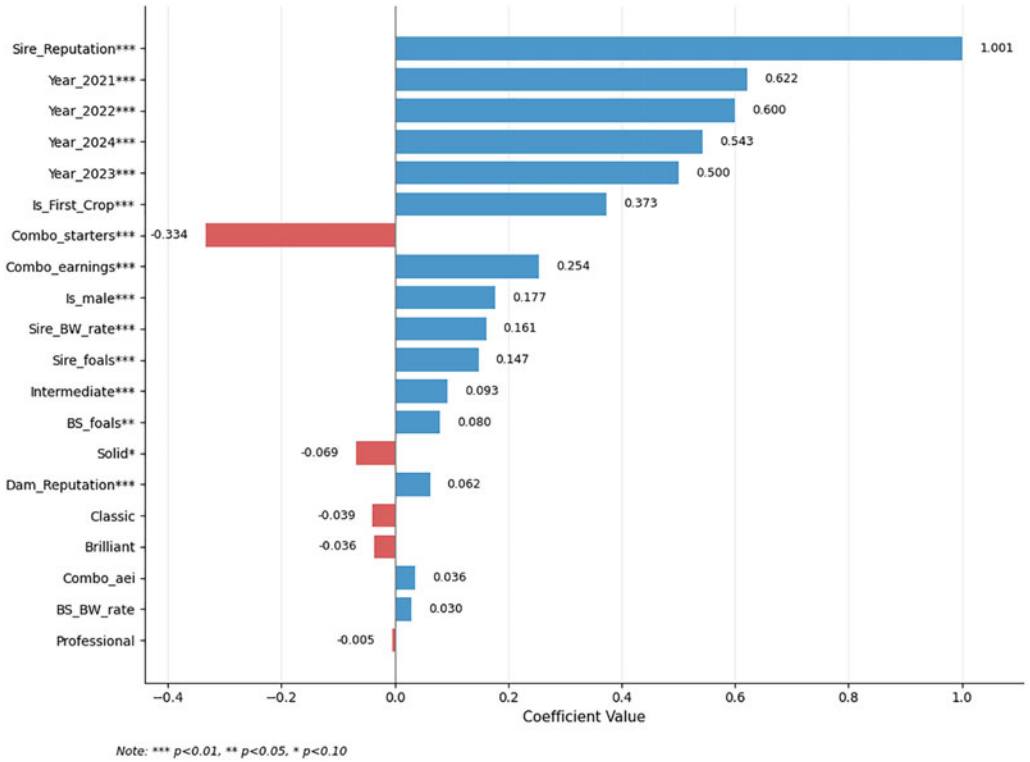

Figure 2. Coefficients from the Ridge model.

sire-related variables explain more price variation than dam characteristics in Thoroughbred markets. The magnitude of this effect supports Neibergs' (2001) argument that buyers use sire reputation as a primary quality signal to reduce information asymmetries, particularly when physical inspection opportunities are limited. Dam Reputation is also positive and statistically significant, with a coefficient of 0.062. We interpret the relative magnitudes of the Sire Reputation and Dam Reputation coefficients with caution. Although the estimated coefficient on Sire Reputation is larger, this should not be construed as evidence that maternal lineage is less economically important in the marketplace. Rather, the difference likely reflects disparities in measurement precision. Stallions sire substantially have more offspring, allowing their market performance to be estimated with lower variance and greater statistical stability. In contrast, individual mares produce far fewer foals over longer time horizons, limiting the observable sample size and increasing estimation noise for Dam Reputation. The smaller Dam coefficient may thus reflect attenuation due to measurement constraints rather than weaker economic relevance.

The estimated coefficient of First-Crop Status (0.373, significant at the 1% level) is also substantial in size. The share of first-crop yearlings is lower in the earliest sessions – 24.20% in Session 1 and 28.57% in Session 2 – than in several later sessions, where it rises above 35% and reaches 41.01% in Session 8. Combined with the greater concentration of sires and dams in Sessions 1 and 2 (Appendix Table A3), this pattern suggests that pricing in the earliest sessions is driven more strongly by established pedigree signals, which may attenuate the observable role of first-crop status in those segments.

Year fixed effects relative to the 2020 baseline are all positive and statistically significant at the 1% level, with 2021 showing the largest coefficient at 0.622, followed by 2022 (0.600), 2024 (0.543), and 2023 (0.500). These coefficients indicate that market conditions throughout the post-2020

period were stronger than in the baseline year. These temporal patterns align with Plant and Stowe's (2013) documentation of cyclical bloodstock markets responding to broader economic conditions. The larger post-2020 coefficients also likely reflect changes in the bloodstock market following the COVID-19 pandemic, including stronger demand and broader shifts in investment conditions.

Among the remaining pedigree and cross-performance variables, Total Earnings for the Sire–Broodmare Sire Cross (0.254), Male Indicator (0.177), Sire Black-type Winner Rate (0.161), and Sire Foals (0.147) all exert positive and statistically significant effects, indicating that both pedigree depth and demonstrated elite performance remain important correlates of auction price. Intermediate Dosage (0.093) is also positive and significant, while Broodmare Sire Foals (0.080) is positive and significant at the 5% level. Starters for the Sire–Broodmare Sire Cross have a relatively large negative coefficient (-0.334, significant at the 1% level). This estimate should be interpreted cautiously, however, because the cross-performance variables are closely related and may capture overlapping dimensions of pedigree history.

Among the other covariates, AEI for the Sire–Broodmare Sire Cross (0.036) and Broodmare Sire Black-type Winner Rate (0.030) are positive but not statistically significant. Several dosage variables have small negative coefficients. Solid Dosage shows a modest negative coefficient (−0.069) and is only marginally significant at the 10% level, whereas Brilliant Dosage (−0.036), Classic Dosage (−0.039), and Professional Dosage (−0.005) are not statistically significant. Overall, these patterns suggest that dosage composition plays a much smaller role in price formation when conditioning on commercial pedigree reputation, year fixed effects, and first-crop status.

4.4. Robustness check and diagnostic analysis

Because the 2020 sale occurred during the COVID-19 pandemic, we examine whether the inclusion of this unusually disrupted year affects our predictive results. Specifically, we test whether the 2020 price distribution differs from other sale years. The Kruskal–Wallis test results indicate that the 2020 price distribution is statistically different from each subsequent sale year (Table A1). This evidence suggests that 2020 represents a distinct pricing environment. Appendix Table A7 reports a robustness check that excludes 2020 observations and re-estimates the Ridge model on the 2021–2024. The sample results achieve an out-of-sample R^2 of 0.5257, with a test RMSE of 0.9947 and a test MAE of 0.8088. These results remain broadly comparable to the full-sample specification, indicating that the paper's main findings are not driven solely by the unusual pricing conditions of the COVID-year sales environment.

We further evaluate temporal stability using a leave-one-year-out cross-validation (LOYO-CV) procedure, in which models are iteratively trained on four years of data and evaluated on the held-out year (Table A2). The LOYO-CV results reveal meaningful variation across folds. In particular, predictive performance declines substantially when 2021 is held out, with out-of-sample R^2 falling from 0.37 in the strongest fold to 0.17 in that case. This pattern suggests that part of the model's explanatory power reflects macroeconomic shifts in addition to cross-sectional pricing relationships. Because year fixed effects are included, these temporal shocks are partially absorbed in estimation. Still, the sensitivity of out-of-sample performance to specific years underscores the importance of caution when interpreting pooled predictive accuracy.

For Ridge model diagnostics, we find that the Ridge model satisfies some, but not all, of the standard regression diagnostics. Normality tests provide mixed evidence. The Jarque–Bera test (statistic = 14.5123, $p = 0.007$) and the Shapiro–Wilk test (statistic = 0.9939, $p = 0.0001$) both reject the null of normally distributed residuals, whereas the Kolmogorov–Smirnov test (statistic = 0.0287, $p = 0.2891$) fails to reject normality. These departures from normality are common in price regressions and are less concerning in a large sample such as ours ($n = 5{,}788$),

Table 4. Model diagnostic test results

Test category	Test	Statistic	*P*-value	Result
Normality	Jarque–Bera	14.5123	0.007	Non-normal
	Shapiro–Wilk	0.9939	0.0001	Non-normal
	Kolmogorov–Smirnov	0.0287	0.2891	Normal
Autocorrelation	Durbin–Watson	2.0523		No correlation
Heteroskedasticity	Breusch–Pagan	36.7068	0.0127	Heteroskedastic
	White	226.0787	0.3052	Homoskedastic

where asymptotic inference remains reliable, particularly when combined with heteroskedasticity-robust standard errors (White, 1980). The Durbin–Watson statistic of 2.0523 is very close to the benchmark value of 2, indicating no meaningful serial correlation in the residuals. Heteroskedasticity tests yield mixed evidence. The Breusch–Pagan test (statistic = 36.7068, $p = 0.0127$) rejects homoskedasticity at conventional significance levels, whereas the White test (statistic = 226.0787, $p = 0.3052$) does not reject constant variance. This discrepancy likely reflects differences in test sensitivity: the Breusch–Pagan test is more responsive to linear forms of heteroskedasticity, whereas the White test is designed to detect more general nonlinear variance patterns.

4.5. Market segmentation analysis

The Keeneland September Yearling Sale is a multi-session auction in which yearlings are strategically placed across sessions based on conformation, pedigree strength, sire power, and recent family sales history. In most years, the sale includes twelve sessions organized across Books 1–5, although the 2021 sale included only eleven sessions. Early sessions feature yearlings with stronger expected commercial appeal, while later sessions include a broader mix of yearlings with more variation in pedigree quality and expected market value (Keeneland Association, 2024). This structured placement reflects meaningful market segmentation. Consistent with this institutional design, Table A3 shows that Sessions 1 and 2 contain fewer unique sires and dams, indicating more concentrated pedigree representation in early sessions. Motivated by this institutional structure, we evaluate model performance at the session level to assess whether predictive accuracy varies systematically across market segments within the sale.

Table 5 shows that model performance varies substantially across Keeneland's 12 sessions, with R^2 ranging from 0.2078 to 0.3725. Session 2 achieves the highest predictive accuracy, followed by Sessions 3 and 9, whereas Sessions 7 and 11 exhibit the weakest fit. The average session-level R^2 of 0.258 (SD = 0.043) indicates moderate predictive power overall but also substantial heterogeneity across sessions. Only one session exceeds an R^2 of 0.30, while six sessions fall below 0.25. This performance gradient does not align neatly with session order. Although several early sessions perform relatively well, predictive accuracy does not decline monotonically as the sale progresses; notably, Session 9 ranks third overall, while Session 7 records the weakest performance despite occurring in the middle of the sale. This pattern suggests that predictability depends less on simple temporal ordering than on the underlying composition of horses within each session.

Table 6 shows that session-level coefficients vary, but most are not statistically significant, likely due to small sample sizes. As a result, these estimates should be interpreted cautiously and not overemphasized. The estimated coefficient on Sire Reputation ranges from 0.5908 in

Table 5. Session-level model performance (ranked by R^2)

Session	R^2	R^2 standard error	RMSE	MAE
2	0.3725	0.0444	0.6022	0.4778
3	0.2861	0.0336	0.7001	0.5219
9	0.2734	0.0435	0.8132	0.6404
1	0.2621	0.0386	0.5793	0.4720
4	0.2621	0.0268	0.7103	0.5574
8	0.2587	0.0328	0.7864	0.6134
10	0.2469	0.0188	0.7420	0.5910
5	0.2469	0.0399	0.7344	0.5749
6	0.2395	0.0520	0.7972	0.6265
12	0.2357	0.0700	0.7406	0.5875
11	0.2099	0.0674	0.7142	0.5756
7	0.2078	0.0240	0.8318	0.6520

Session 3 to 0.2081 in Session 12, a nearly threefold difference. This pattern suggests that buyers in earlier sessions incorporate stallion quality more systematically into their valuations. Regarding the first-crop status, the estimated coefficient is strongly positive in Session 3 (0.3769) and remains sizable in Session 12 (0.3363), but it is negative in Sessions 1, 2, and 4. This variation indicates that the speculative value associated with unproven sires is not uniform across market segments.

Table 6 also reinforces the segmentation evident in Table 5. Our model appears to provide its greatest practical value in Sessions 2 and 3, where predictive fit is strongest and pedigree signals are more systematically related to price. By contrast, in Sessions 7 and 11, where R^2 is 0.2078 and 0.2099, respectively, pedigree-based valuation appears to offer less information. Buyers in these sessions may therefore need to rely more heavily on physical inspection and subjective assessment. More broadly, these patterns are consistent with differences in price-discovery efficiency across market segments. Higher-quality sessions exhibit stronger alignment between observable pedigree signals and prices, creating more efficient price discovery (Chezum and Wimmer, 1997). Lower-quality sessions feature more casual participants and greater information asymmetries, leading to noisier pricing. This aligns with findings from other agricultural markets where product heterogeneity reduces pricing efficiency (Wang et al., 2025).

5. Discussion

Our findings show that machine learning models can capture a meaningful proportion of variation in Thoroughbred yearling auction prices (out-of-sample R^2 around 0.54). This level of explanatory power indicates that pedigree information, while incomplete, remains important for yearling auction pricing. Model performance varies substantially across Keeneland's session structure. Predictive accuracy peaks in Sessions 2 and 3 ($R^2 \approx 0.3725$ and 0.2861, respectively), where observable pedigree signals appear to align more closely with prices. Session 9 also performs relatively well ($R^2 = 0.2734$), indicating that strong model fit is not confined exclusively to the earliest books of the sale. Predictive performance declines to approximately 0.21 in Sessions 7 and 11, suggesting that in weaker-fit sessions, pedigree information captures a smaller share of price variation.

Table 6. Session-level ridge regression coefficients

	Sessions											
	1	2	3	4	5	6	7	8	9	10	11	12
Dam reputation	0.0156	0.0212	0.0252	0.0223	0.0335	0.0577	0.0512	0.0478	0.0697	0.0606	0.0333	0.0223
Sire reputation	0.3166	0.4419	0.5908	0.3894	0.5373	0.3950	0.4243	0.5010	0.3691	0.2820	0.2325	0.2081
Year 2021	−0.0432	−0.1059	0.4649	0.0702	0.1138	0.3146	0.6564	0.7596	0.6644	0.5677	0.6497	0
Year 2022	0.0696	0.0739	0.5545	0.0306	0.3292	0.3934	0.6127	0.7022	0.4866	0.4273	0.6669	0.8420
Year 2023	0.0563	−0.0061	0.4122	0.0658	0.2189	0.3143	0.4428	0.6594	0.6530	0.4083	0.6335	0.8732
Year 2024	0.0764	0.1758	0.6008	0.1295	0.2648	0.3622	0.4807	0.6933	0.6621	0.3367	0.4910	0.9349
Is First Crop	−0.0870	−0.0931	0.3769	−0.1429	0.1378	0.1793	0.2388	0.1648	0.0717	0.0791	0.1165	0.3363
Is Male	0.1048	0.0323	0.0245	0.0736	0.0782	0.0537	0.3291	0.1341	0.1265	0.2086	0.1432	0.1438
Brilliant	−0.0557	0.0228	−0.1192	−0.0494	−0.0805	0.0361	−0.0395	−0.0028	−0.0739	−0.0623	0.0442	−0.0974
Classic	−0.0696	−0.1557	0.0637	−0.0477	−0.0816	−0.1181	−0.0526	0.0777	0.0567	−0.0361	0.0277	0.1715
Combo AEI	0.0706	0.0454	−0.0514	0.0650	−0.0416	−0.1053	0.1289	−0.0296	0.221	0.0961	0.0653	−0.0823
Combo starters	0.0208	0.0125	−0.1708	−0.1708	−0.1348	−0.2761	0.0599	−0.3273	−0.0486	−0.4667	−0.2624	−0.2805
Combo earnings	0.0259	−0.0771	0.2281	0.0196	0.1689	0.2565	−0.1226	0.2776	−0.0093	0.2355	0.1527	0.4139
Intermediate	0.0184	0.0158	0.0835	0.0426	0.0181	0.0282	0.0523	0.0272	0.0975	0.1917	0.0865	0.1032
Professional	−0.0011	−0.0017	−0.0430	−0.0006	0.0067	0.0004	0.0222	0.0042	0.0257	−0.0066	0.0096	0.0180
Solid	−0.0639	−0.0935	−0.1980	−0.1001	−0.0064	0.0806	−0.0040	−0.1501	−0.0319	0.1190	0.0371	0.0446
BS BW rate	−0.0242	0.0481	−0.0070	0.0189	0.0127	0.0340	0.0410	0.0546	0.0212	−0.0843	−0.0056	0.0304
BS foals	−0.0266	0.1160	−0.0336	0.0413	0.0341	−0.0438	0.0267	−0.0537	−0.0159	0.1855	0.1521	−0.0157
Sire BW rate	0.0475	0.2038	0.0698	0.1229	0.1010	0.1891	0.1176	−0.0272	0.0975	0.0556	0.1978	0.1316
Sire foals	0.0510	0.0197	0.0306	−0.0617	−0.0627	−0.1015	−0.0321	−0.0103	−0.0712	0.0803	0.0559	−0.0192

(*Continued*)

Table 6. (*Continued*)

	Sessions											
	1	2	3	4	5	6	7	8	9	10	11	12
Model evaluations												
Observations	219	224	438	477	576	569	622	634	636	654	390	349
Features	20	20	20	20	20	20	20	20	20	20	20	20
R^2	0.2621	0.3725	0.2861	0.2621	0.2469	0.2395	0.2078	0.2587	0.2734	0.2469	0.2099	0.2357
Adjusted R^2	0.1876	0.3107	0.2519	0.2298	0.2198	0.2117	0.1815	0.2346	0.2497	0.2231	0.1671	0.1890
RMSE	0.5793	0.6022	0.7001	0.7103	0.7344	0.7972	0.8318	0.7864	0.8132	0.7420	0.7142	0.7406
MAE	0.4720	0.4778	0.5219	0.5574	0.5749	0.6265	0.6520	0.6134	0.6404	0.5910	0.5756	0.5875
Regularization (α)	32.3746	15.9986	0.7197	42.9193	5.1795	4.4984	2.5595	0.7197	2.5595	2.223	0.7197	0.6251

Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors computed using 1,000 bootstrap samples.

The dominance of reputation-based variables provides important insight into market information under asymmetric conditions. Sire Reputation emerges as the most influential predictor, with the largest positive coefficient in the Ridge model. This finding echoes Hansen and Stowe (2018) and Neibergs (2001), who emphasize the centrality of sire-related signals in reducing buyer uncertainty. However, Dam Reputation remains economically meaningful despite its smaller estimated coefficient. This smaller magnitude likely reflects greater measurement noise, as dams have far fewer observable offspring than sires. Notably, first-crop status also remains economically important, confirming that novelty and speculation drive buyers' willingness to pay for unproven sires.

Several limitations warrant consideration. First, the Sire Reputation index is constructed from historical auction outcomes, which raises potential concerns regarding circularity and endogeneity. Because past prices may embed persistent market perceptions and unobserved quality signals, the estimated coefficients should not be interpreted as causal effects. Rather, the model identifies predictive associations conditional on observed information. Second, the analysis is based on a single major auction venue and a specific time window, which limits external validity. Pricing dynamics may differ across sales formats (e.g., select vs. open sales), geographic regions, buyer composition, or broader macroeconomic conditions. While the Keeneland September Sale represents the largest and most influential yearling market, results may not fully generalize to other platforms such as Fasig-Tipton or Ocala Breeders' Sales Company (OBS), where market structure, buyer compositions, and information environments differ. Future research should extend the framework to multiple auction venues and longer time horizons to assess cross-market robustness. Comparative analysis across sales formats and macroeconomic regimes would help disentangle venue-specific effects from broader structural pricing mechanisms. Such multi-market validation would strengthen external generalizability and provide a more comprehensive understanding of Thoroughbred price formation. Third, some session-level comparisons are based on relatively small subsamples, which may increase estimation variance and reduce precision. Fourth, approximately 46% of the variation in Thoroughbred yearling auction prices remains unexplained. This reflects the limits of catalog data in capturing key determinants of value, particularly physical conformation, soundness, and subjective assessments. The weak and inconsistent predictive power in later and weaker-fit sessions reinforces the conclusion that pedigree-based models provide less guidance where quality converges and idiosyncratic factors dominate. Future research could expand predictive accuracy by integrating multimodal data sources, such as computer vision analysis of yearling conformation, natural language processing of catalog narratives, and veterinary or behavioral records.

Our findings have several practical and policy implications for market participants. For sellers, these models can inform reserve pricing and consignment strategy, especially in elite sessions where pedigree signals are most predictive. For buyers, our models provide data-driven benchmarks to complement subjective judgments, reducing uncertainty in a market characterized by information asymmetries. For breeders, the strong role of sire and reputation-based signals underscores the economic value of long-term investment in pedigree quality and performance visibility, as reputation appears to function as a central valuation anchor in the yearling market. For consignors, the variation in predictability across sessions suggests that sale placement strategy may materially influence pricing outcomes, particularly in higher-tier sessions where observable information is more systematically capitalized into bids. For auction organizers, the results highlight the importance of transparent and standardized information disclosure, as clearer pedigree and performance metrics may enhance price discovery and reduce uncertainty. More broadly, strengthening reporting practices and facilitating consistent access to reputational data may improve market efficiency by allowing buyers to more effectively process available signals when forming bids.

6. Conclusion

Our study demonstrates that machine learning methods, particularly Ridge regression, can effectively model Thoroughbred yearling prices at the Keeneland September Sale. Our final specification explains approximately 54% of out-of-sample log-price variance across 5,788 yearlings sold between 2020 and 2024, placing our results well within the range of prior hedonic studies of bloodstock markets (Plant and Stowe, 2013; Robert and Stowe, 2016).

While our analysis focuses on Thoroughbred yearling sales, the methodological framework has broader relevance. Many agricultural markets use auctions to allocate heterogeneous assets under uncertainty. Markets for cattle, timber, fish, and online livestock, for example, all involve variation in product quality and substantial information asymmetry between buyers and sellers. The framework we develop offers a replicable template for future studies on asset valuation under uncertainty, where both predictive performance and economic interpretability are essential.

Future research should integrate multimodal and non-traditional data sources. Computer vision applied to conformation images, natural language processing of catalog descriptions or veterinary records, and real-time bidding dynamics represent promising extensions. Incorporating such features may improve predictive accuracy and yield a more comprehensive understanding of how both observable and latent signals shape price formation. As institutional investors and professional buyers become more active in bloodstock markets, reliable data-driven valuation tools will become increasingly important. These tools can help buyers evaluate yearlings more systematically and allocate capital more efficiently.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/aae.2026.10050.

Data availability statement. The data used in this study was compiled from two primary sources. Sale transaction records (2020–2024) were collected from publicly available archives on Keeneland's official website (www.keeneland.com), including sale prices, hip numbers, sex, color, and session placement[2]. Pedigree and performance metrics were obtained from Equineline.com Free 5-Cross Pedigree reports, operated by The Jockey Club Information Systems (TJCIS)[3]. The compiled dataset integrating these sources is available from the corresponding author upon reasonable request. Researchers seeking to replicate this study should note that while Keeneland sale results are publicly accessible, Equineline data usage may be subject to TJCIS terms of service. The analysis code and processed dataset are available upon request to the corresponding author.

Author contribution. Conceptualization, J.C. and Y.Y.; Methodology, T.N., T.P., and Y.Y.; Formal Analysis, T.N., T.P. and Y.Y.; Data Curation, T.N. and T.P.; Literature Review, T.M.; Writing – Original Draft Preparation, Y.Y.; Writing – Review and Editing, T.N. and Y.Y.; Supervision, J.C. and Y.Y.; Project Administration, J.C. and Y.Y.

Financial support. This study was conducted without external funding from public, commercial, or not-for-profit sources.

Competing interests. The authors have no conflicts of interest

Declaration of AI use. Artificial intelligence tools (e.g., ChatGPT) were used for minor editorial assistance, such as checking grammar, spelling, and typographical errors. No AI tools were employed in the design of the study, data analysis, interpretation of results, or drafting of substantive content. The authors take full responsibility for the content of the manuscript.

References

Alani, H., and C. Brewster. "Metrics for Ranking Ontologies." In Proceedings of the International Semantic Web Conference (ISWC), 2006.

American Horse Council. "*Results From the 2023 National Equine Economic Impact Study Released*."American Horse Council, 2023. Internet site: https://horsecouncil.org/project/results-from-the-2023-national-equine-economic-impact-study-released/ (Accessed April 2026).

Athey, S., and G. Imbens. "Machine learning methods that economists should know about." *Annual Review of Economics* **11**, (2019):685–725. https://doi.org/10.1146/annurev-economics-080217-053433

[2]Keeneland Association. Sale Transaction Records. 2020–2024.

[3]The Jockey Club Information Systems. Equineline.com Free 5-Cross Pedigree Reports.

Box, G.E.P., and D.R. Cox. "An analysis of transformations." *Journal of the Royal Statistical Society: Series B (Methodological)* **26**,(1964):211–43. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Breusch, T.S., and A.R. Pagan. "A simple test for heteroscedasticity and random coefficient variation." *Econometrica* **47**,5(1979): 1287–94.

Buzby, J.C., and E.L. Jessup. "The relative impact of macroeconomic and yearling-specific variables in determining thoroughbred yearling price." *Applied Economics* **26**,1(1994):1–8.

Calil, Y.C.D., L.A. Ribera, D.P. Anderson, and W. Koury-Filho. "Pure-Bred Nellore prices in Brazil: Morphological, genetic, physical, and market factors in auctions." *Journal of Agricultural and Resource Economics* **47**,3(2022):529–43.

Chezum, B., and B. Wimmer. "Roses or lemons: Adverse selection in the market for thoroughbred yearlings." *The Review of Economics and Statistics* **79**,3(1997):521–30.

Cook, R.D. "Detection of influential observation in linear regression." *Technometrics* **19**,1(1977):15–8.

Dormann, C.F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J.R. García Marquéz, et al. "Collinearity: A review of methods to deal with it and a simulation study evaluating their performance." *Ecography* **36**,(2013):27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Durbin, J., and G.S. Watson. "Testing for serial correlation in least squares regression." *Biometrika* **3**,4(1950):409–28. https://doi.org/10.2307/2332391

Godfrey, L.G. "Testing against General ARMA Errors when Regressors Include Lagged Dependent Variables." *Econometrica* **46**,6(1978):1293–301.

Hansen, C.R., and C.J. Stowe. "Determinants of weanling thoroughbred auction prices." *Journal of Agricultural and Applied Economics* **50**,1(2018):48–63. https://doi.org/10.1017/aae.2017.19

Haque, M.M., M.A.I. Talukder, S.S. Saha, and M.S. Uddin. "A machine learning-based price prediction for cows." *American Journal of Agricultural Science, Engineering, and Technology* **5**,1(2021):29–40. https://doi.org/10.5281/zenodo.4606334

Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer, 2009.

Jarque, C.M., and K. Anil. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals." *Economics Letters* **6**(1980):255–9. https://doi.org/10.1016/0165-1765(80)90024-5

Keeneland Association. "*Keeneland September Yearling Sale Becomes Highest-Grossing Auction in Keeneland History*. Keeneland Media." 2024. Internet site: https://www.keeneland.com/media/news/keeneland-september-yearling-sale-becomes-highest-grossing-auction-keeneland-history

Kibler, M.L., and J.M. Thompson. "Price determinants of stock-type horses sold at public online auctions." *Journal of Agricultural and Applied Economics* **52**,4(2020): 596–612. https://doi.org/10.1017/aae.2020.20

Kim, J.S., S.D. Mitchell and L.C. Wang. "Hedonic pricing and the role of stud fees in the market for thoroughbred yearlings in Australia." *Australian Journal of Agricultural and Resource Economics* **63**,3(2019):439–71. https://doi.org/10.22004/ag.econ.333838

Maynard, L.J., and K.M. Stoeppel. "Hedonic price analysis of thoroughbred broodmares in foal." *Journal of Agribusiness* **25**, 2(2007):181–95. https://doi.org/10.22004/ag.econ.62295

Milgrom, Paul R., and R.J. Weber. "A theory of auctions and competitive bidding." *Econometrica* **50**,5(1982):1089–122. https://doi.org/10.2307/1911865

Mouncey, R.R., P. Alarcon, and K.L. Verheyen. "Determinants of thoroughbred yearling sales price in the UK." *Veterinary Record Open* **11**,(2024):e81. https://doi.org/10.1002/vro2.81

Mullainathan, S., and J. Spiess. "Machine learning: An applied econometric approach." *Journal of Economic Perspectives* **31**,2(2017): 87–106. https://doi.org/10.1257/jep.31.2.87

Neibergs, J.S. "A hedonic price analysis of thoroughbred broodmare characteristics." *Journal of Agribusiness* **17**,2(2001):299–314.

Parsons, C., and I. Smith. "The price of thoroughbred yearlings in Britain." *Journal of Sports Economics* **9**,1(2007):43–66. https://doi.org/10.1177/1527002506296550

Plant, E.J., and C.J. Stowe. "The price of disclosure in the thoroughbred yearling market." *Journal of Agricultural and Applied Economics* **45**,2(2013):243–57. https://doi.org/10.22004/ag.econ.149125

Robert, M., and C.J. Stowe. "Ready to run: Price determinants of thoroughbreds from 2-year-olds in training sales." *Applied Economics* **48**,48(2016):4690–7.

Rogers, S.G., T.C. Schroeder, G.T. Tonsor, and B.K. Coffey. "Describing variation in formula base prices for U.S. fed cattle: A hedonic approach." *Journal of Agricultural and Applied Economics* **55**,1(2023):117–32.

Schroeder, G.F., T.R. Koontz, and R.G. Wagner. "Factors affecting prices in auction and private treaty markets for feeder cattle." *Journal of Agricultural and Resource Economics* **13**,2(1988):225–35.

Schroeder, T.C., J.R. Mintert, F. Brazle, and O.C. Grunewald. "Factors affecting feeder cattle price differentials." *Western Journal of Agricultural Economics* **13**,1(1988):1–11.

Schroeder, T.C., B.K. Coffey, and G.T. Tonsor. "Hedonic modeling to facilitate price reporting and fed cattle market transparency." *Applied Economic Perspectives and Policy* **45**,3(2023):1716–33.

Schulz, L.L., K.C. Dhuyvetter, and B.E. Doran. "Factors affecting preconditioned calf price premiums: does potential buyer competition and seller reputation matter?" *Journal of Agricultural and Resource Economics* **40**,2(2015):220–41.

Shapiro, S.S., and M.B. Wilk. "An analysis of variance test for normality (complete samples)." *Biometrika* **3**,4(1965):591–611. https://doi.org/10.2307/2333709.

Varian, H.R. "Big data: New tricks for econometrics." *Journal of Economic Perspectives* **28**,2(2014):3–28. https://doi.org/10.1257/jep.28.2.3.

Vickner, S.S., and S.I. Koch. "Hedonic pricing, information, and the market for thoroughbred yearlings." *Journal of Agribusiness* **19**,2(2001):173–89. https://doi.org/10.22004/ag.econ.14693

Wang, Z., M.-K. Kim, and H. Tejeda. "Unraveling hidden patterns in fed cattle negotiated cash prices using machine learning." *Journal of Agricultural and Applied Economics* **57**,4(2025):563–581. https://doi.org/10.1017/aae.2025.10009

White, H. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica* **48**,4(1980):817–38.

Yeo, In-Kwon, and R.A. Johnson. "A new family of power transformations to improve normality or symmetry." *Biometrika* **87**,4(2000):954–59. http://www.jstor.org/stable/2673623.

Zapata, S.D., and D. Anderson. "Benchmarking analysis of cattle auction prices." *Journal of Agricultural and Applied Economics* **57**,3(2025):393–410. https://doi.org/10.1017/aae.2025.10011

Zimmerman, L.C., T.C. Schroeder, K.C. Dhuyvetter, K.C. Olson, G.L. Stokka, J.T. Seeger, and D.M. Grotelueschen. "The effect of value-added management on calf prices at superior livestock auction video markets." *Journal of Agricultural and Resource Economics* **37**,1(2012):1–18.

Cite this article: Yang, Y., J. Clarke, T. Mandal, T. Nguyen, and T. Pham (2026). "Predicting Thoroughbred Yearling Auction Prices with Machine Learning: Evidence from the Keeneland September Sale." *Journal of Agricultural and Applied Economics* 1–19. https://doi.org/10.1017/aae.2026.10050