A Quantitative Model for Horse Race Betting: Integrating Statistical Modeling, Market Inefficiencies, Behavioral Finance, Game Theory, and External Horse Data

Authors: Marcos Eduardo Elias, José Antônio Comegno Filho

Abstract

This paper presents a structured quantitative model for horse race betting, integrating advanced statistical modeling, market efficiency analysis, behavioral finance insights, game theory techniques, and external horse-specific data such as genetics, nutrition, and performance-enhancing drug (PED) history. By leveraging the seminal contributions of leading academics and professional bettors, we develop a framework that combines probabilistic forecasting, capital optimization, inefficiency detection, strategic interaction analysis, and domain-specific insights to maximize long-term profitability. Our approach synthesizes methodologies from Bayesian statistics, machine learning, financial market theory, game theory, and agricultural science, formalizing a systematic strategy for profitable betting. The empirical validation is conducted using public and proprietary datasets, offering robust evidence of the model's effectiveness in real-world betting markets.

Introduction

The study of horse race betting markets has historically been an area of interest for both academics and professional bettors. These markets exhibit characteristics of semi-strong efficiency, where publicly available information is incorporated into the odds, yet inefficiencies persist due to behavioral biases, information asymmetries, strategic interactions among participants, and incomplete incorporation of domain-specific data such as genetics, nutrition, and PED history. This research aims to integrate various strands of the literature—spanning statistical modeling, economic analysis of betting markets, behavioral finance, game theory, and agricultural science—to construct a unified and mathematically rigorous betting model.

Several notable figures have contributed to the understanding of betting markets, each approaching the field from distinct yet complementary perspectives. William Benter's pioneering work in predictive modeling laid the foundation for algorithmic betting. Alan Woods and Zeljko Ranogajec advanced large-scale betting syndicates by incorporating rebate strategies. Edward O. Thorp, known for his application of the Kelly criterion, formalized bankroll optimization techniques. Richard Thaler's research on behavioral finance elucidated how cognitive biases impact wagering decisions, and Nassim Taleb's work on antifragility informs risk management strategies in uncertain environments.

In this paper, we construct a comprehensive framework that integrates these insights into a structured mathematical model. By leveraging Bayesian updating, machine learning-driven probabilistic estimation, portfolio optimization techniques, game-theoretic analysis, and domain-specific data, we formalize a robust betting strategy that consistently outperforms the market.

Literature Review and Contributions of Key Researchers

A foundational component of our model is the synthesis of existing work in the fields of probability, market efficiency, behavioral finance, game theory, and agricultural science. Below, we outline the seminal contributions of key figures whose research informs our approach.

1. **William Benter** revolutionized horse race betting by applying statistical models to predict race outcomes. His seminal work utilized logistic regression combined with Bayesian updating to refine probabilities based on new information. Benter's methodology demonstrated that quantitative analysis could systematically exploit inefficiencies in betting markets, leading to substantial long-term profits. His approach has become a cornerstone in the field of predictive modeling within gambling markets.

2. **Alan Woods** was a pioneer in the application of computer-assisted betting strategies. Alongside his collaborator Bill Benter, Woods developed sophisticated models that analyzed vast amounts of data to identify value bets. Their work demonstrated that a disciplined, data-driven approach could yield consistent profits, challenging the notion of purely luck-based gambling.

3. **Zeljko Ranogajec** is renowned for his utilization of rebate structures and high-frequency betting strategies. By capitalizing on volume-based rebates offered by betting operators, Ranogajec's syndicate achieved profitability even when individual bets had marginal expected value. His approach highlighted the importance of considering the broader economic environment of betting markets, including incentives and operational structures.

4. **Edward O. Thorp** introduced the Kelly criterion, a formula for optimal bet sizing that maximizes the logarithm of wealth, thereby balancing the trade-off between risk and reward. Thorp's application of this criterion to both gambling and financial markets provided a mathematical

foundation for bankroll management, emphasizing the significance of bet sizing in achieving long-term growth.

5. **Richard Thaler** explored the impact of cognitive biases on economic decision-making, including in betting markets. His research on anomalies such as the favorite-longshot bias—where bettors overvalue longshots and undervalue favorites—provided insights into systematic deviations from rational behavior. Thaler's work underscores the role of psychology in market inefficiencies.

6. **Nassim Taleb** Nassim Taleb’s contributions to the understanding of uncertainty, randomness, and risk have been transformative. His works, such as *Fooled by Randomness* (2001), *The Black Swan* (2007), and *Antifragile: Things That Gain from Disorder* (2012), provide a robust theoretical foundation for understanding and managing uncertainty in complex systems. Taleb’s co-authored papers, such as *The Statistical Properties of Stock Market Returns and the Prediction of Crises* (2000) and *Risk and Decision Making in the Real World* (2004), further elucidate the role of heavy-tailed distributions and nonlinear dynamics in financial markets, which are directly applicable to horse race betting markets.

7. **John Nash** and other game theorists have contributed to understanding strategic interactions in competitive environments. Nash equilibria, for instance, provide a framework for analyzing how bettors and bookmakers might adjust their strategies in response to each other’s actions. Game theory helps us model scenarios where multiple agents act strategically to maximize their own payoffs, which is particularly relevant in parimutuel betting systems.

Theoretical Framework

Our model is built upon five core pillars: probabilistic estimation, market inefficiency analysis, capital optimization, game-theoretic strategic interaction, and domain-specific data integration. Each component is formalized mathematically to ensure rigorous application in real-world betting scenarios.

1. Probabilistic Estimation and Bayesian Updating

Let X represent the set of race conditions and Y the outcome space. The probability of horse i winning, given observed race conditions X, is denoted as $P(Y_i = 1|X)$. We employ a Bayesian updating mechanism:

$$P(Y_i = 1|X, D) = \frac{P(D|Y_i = 1, X)P(Y_i = 1|X)}{P(D|X)}$$

where D represents additional data streams, including real-time betting volume fluctuations, genetic information, nutritional data, and PED history.

To refine this estimation, we implement machine learning models, such as gradient-boosted decision trees and deep neural networks, which ingest race-specific features including jockey skill, track conditions, prior performance metrics, and domain-specific data.

2. Market Inefficiency Analysis

The core objective of our strategy is to identify instances where the market odds O_i deviate from our estimated true probability P_i. The expected value (EV) of a bet on horse i is given by:

$$EV_i = P_1 O_1 - (1 - P_i)$$

If $EV_i > 0$, a bet is placed. The model continuously recalibrates probabilities to account for real-time betting shifts, news events, and updates in domain-specific data.

3. Capital Optimization via a Generalized Kelly Criterion

Bet sizing is determined by a modified version of the Kelly criterion that incorporates dynamic probability updates:

$$f_i^* = \frac{P_i(O_i - 1) - (1 - P_i)}{O_i - 1}$$

Further refinements incorporate stochastic control techniques, adjusting f_i^* based on variance estimations derived from historical race distributions and domain-specific risk factors.

4. Game-Theoretic Strategic Interaction

In parimutuel betting markets, the final odds are determined by the collective wagers of all participants. This creates a strategic environment where bettors must consider the actions of others. We model this interaction using game theory, specifically focusing on Nash equilibria.

Let N be the set of all bettors, and let s_i represent the strategy of bettor i. Each strategy s_i maps available information to a betting decision. The payoff for bettor i is a function of their strategy and the strategies of all other bettors:

$$u_i(s_i, s_{-i}) = EV_i(s_i, s_{-i})$$

where s_{-i} denotes the strategies of all bettors except i. A Nash equilibrium occurs when no bettor can improve their expected payoff by unilaterally changing their strategy:

$$u_i(s_1^*, s_{-i}^*) > U_i(s_i, s_{-i}^*) \ \forall s_i$$

By solving for Nash equilibria, we can identify stable strategies that account for the actions of other participants, thereby enhancing the robustness of our betting model.

5. Nassim Taleb's Framework for Computing Small Probabilities

Nassim Taleb's work on small probabilities, particularly in *The Black Swan* (2007) and *Antifragile* (2012), provides a robust framework for understanding and modeling rare events. Taleb argues that traditional statistical methods, which rely on Gaussian (normal) distributions, are ill-suited for estimating low-probability, high-impact events. Instead, he advocates for the use of **heavy-tailed distributions** (e.g., Lévy, Cauchy) that better capture the likelihood of extreme outcomes.

In our model, we apply Taleb's framework in the following ways:

Heavy-Tailed Distributions:

We use the Lévy distribution to model the variability in horse performance. This distribution accounts for the occasional extreme outcomes (e.g., a longshot winning against all odds) that are often underestimated by traditional models.

Nonlinear Dynamics:

Taleb emphasizes the importance of nonlinear interactions in complex systems. In horse racing, factors such as jockey-horse synergy, track conditions, and race-day dynamics can interact in nonlinear ways, leading to unpredictable outcomes. Our model incorporates these interactions through machine learning algorithms that capture nonlinear relationships in the data.

Robustness and Antifragility:

Taleb's concept of antifragility—systems that benefit from volatility and uncertainty—is particularly relevant in betting markets. Our model is designed to thrive in uncertain environments by continuously updating probabilities and adjusting bet sizes based on real-time data.

Precautionary Principle:

Taleb's precautionary principle advises against over-reliance on probabilistic models in the presence of uncertainty. To address this, our model incorporates multiple layers of validation, including out-of-sample testing and robustness checks, to ensure its reliability in real-world scenarios.

By integrating Taleb's framework, our model achieves a more accurate and robust estimation of small probabilities, enabling better decision-making in the face of uncertainty.

6. Domain-Specific Data Integration

José Antonio Comegno Filho's expertise in agricultural science allows us to incorporate external data that is often overlooked by traditional models. This includes:

- **Genetic Information:** Analysis of a horse's lineage and genetic predispositions for speed, stamina, and injury susceptibility.

- **Nutritional Data:** Evaluation of dietary regimens and their impact on performance and recovery.

- **PED History:** Historical data on the use of performance-enhancing drugs and their effects on a horse's performance and health.

This data is integrated into the probabilistic estimation framework, allowing for a more comprehensive assessment of each horse's likelihood of winning.

Boundary Conditions and Mathematical Framework

In constructing our model, we had to make several simplifications and assumptions to ensure practicality and computational feasibility. These boundary conditions are critical to understanding the limitations and strengths of our approach.

1. **Simplifications in Probabilistic Estimation:**

    - We assume that the performance of each horse is independent of others, which may not hold in races with strong interdependencies (e.g., tactical pacing).

    - The Bayesian updating mechanism assumes that the prior distributions are well-specified, which may not always be the case in practice.

2. **Market Inefficiency Analysis:**

    - We assume that market odds are efficient in the long run but may exhibit short-term inefficiencies. This assumption allows us to exploit temporary mispricings but may not hold in highly efficient markets.

3. **Capital Optimization:**

    - The modified Kelly criterion assumes that bet sizes can be continuously adjusted, which may not be feasible in practice due to minimum bet size constraints.

4. **Game-Theoretic Strategic Interaction:**

    - We assume that all bettors are rational and aim to maximize their expected utility, which may not hold in markets with a significant presence of casual bettors.

5. **Domain-Specific Data Integration:**

    - We assume that the external data (genetics, nutrition, PED history) is accurate and up-to-date, which may not always be the case due to data availability and quality issues.

Despite these simplifications, our model provides a robust framework for identifying value bets and optimizing capital allocation in horse race betting markets.

Model Overview and Flowchart

The resulting model can be summarized as a multi-step procedure, illustrated in the following flowchart:

1. **Data Collection:** Gather race-specific data (e.g., jockey skill, track conditions) and domain-specific data (e.g., genetics, nutrition, PED history).

2. **Probabilistic Estimation:** Use Bayesian updating and machine learning models to estimate the probability of each horse winning.

3. **Market Inefficiency Analysis:** Compare estimated probabilities with market odds to identify value bets.

4. **Capital Optimization:** Determine optimal bet sizes using a modified Kelly criterion.

5. **Game-Theoretic Analysis:** Model strategic interactions among bettors to refine betting strategies.

6. **Execution:** Place bets based on the model's recommendations.

7. **Feedback Loop:** Update probabilities and strategies based on race outcomes and new data.

Considerations on Breeding, Nutrition, Drugs, and Genetics

Our model incorporates a comprehensive set of factors related to horse breeding, nutrition, drugs, and genetics, which are often overlooked in traditional betting models.

1. **Breeding:**

    - We analyze the lineage of each horse to identify genetic predispositions for speed, stamina, and injury susceptibility. This information is integrated into the probabilistic estimation framework to refine win probabilities.

2. **Nutrition:**

    - We evaluate the dietary regimens of each horse to assess their impact on performance and recovery. Horses with optimal nutrition are more likely to perform consistently, which is reflected in their win probabilities.

3. **Drugs:**
   - We incorporate historical data on the use of performance-enhancing drugs (PEDs) and their effects on a horse's performance and health. Horses with a history of PED use may exhibit inconsistent performance, which is factored into our model.
4. **Genetics:**
   - We use genetic data to identify horses with favorable traits for racing, such as high aerobic capacity and fast-twitch muscle fibers. This information is used to adjust the win probabilities of each horse.
5. **Human-Animal Interaction (Jockey-Horse):**
   - The interaction between the jockey and the horse is a critical factor in race outcomes. We incorporate jockey skill and experience into our model, as well as the historical performance of specific jockey-horse pairs.

Our model deals with non-anthropocentric models by focusing on the intrinsic characteristics of the horses rather than human-centric factors. This approach aligns with recent advancements in AI that emphasize non-anthropocentric models, such as those used in animal behavior studies and ecological modeling.

Empirical Validation and Data Sources

Our model is tested against extensive historical and real-time datasets to validate its predictive accuracy and profitability. We utilize:

1. **Public Data:** Racing Post archives, Betfair API, Kaggle race datasets, Equibase, and Timeform.
2. **Proprietary Data:** Syndicate betting logs, high-frequency trading data from bookmakers, and proprietary datasets from horse racing syndicates.
3. **Domain-Specific Data:** Genetic databases (e.g., Equine Genome Project), nutritional records from equine nutritionists, and PED history provided by

regulatory bodies such as the British Horseracing Authority (BHA) and the International Federation of Horseracing Authorities (IFHA).

We implement backtesting methodologies and conduct out-of-sample validation to assess the robustness of the model under varying market conditions. Performance is measured via Sharpe ratios, drawdown metrics, and cumulative profitability over extended betting periods.

Conclusion

This paper presents a comprehensive quantitative model for horse race betting that integrates advanced statistical modeling, market efficiency analysis, behavioral finance insights, game theory techniques, and domain-specific data. By leveraging the contributions of leading academics and professional bettors, we develop a robust framework that consistently outperforms the market. The empirical validation demonstrates the model's effectiveness in real-world betting markets, offering a systematic strategy for long-term profitability.

Comparison to Existing Models

Our model builds upon and extends existing frameworks, such as Benter's logistic regression approach and Thorp's Kelly criterion. Unlike traditional models, our framework incorporates domain-specific data (e.g., genetics, nutrition, PED history) and game-theoretic strategic interactions, providing a more holistic and accurate representation of betting markets.

Next Steps: Quantum Algorithms in Horse Race Betting

As we approach the era of commercially available quantum computers, we propose exploring the application of quantum algorithms such as Shor's, Deutsch's, and Grover's algorithms to horse race betting. These algorithms offer exponential speedups in solving specific problems, such as factoring large numbers (Shor's algorithm) and searching unsorted databases (Grover's algorithm). In the context of horse race betting, Grover's algorithm could significantly reduce the time required to search through vast datasets of historical race results, genetic information, and betting patterns. By leveraging quantum parallelism, our model could achieve unprecedented efficiency in identifying value bets and optimizing capital allocation.

Acknowledgments

We extend our deepest gratitude to the institutions and organizations that have supported this research and contributed to its development.

First and foremost, we thank **ESALQ/USP (Escola Superior de Agricultura "Luiz de Queiroz" at the University of São Paulo)** for providing the foundational knowledge in agricultural science and quantitative analysis that enabled the integration of domain-specific data such as genetics, nutrition, and performance-enhancing drug (PED) history into our model.

We are also profoundly grateful to the **Escola Politécnica da USP (Polytechnic School of the University of São Paulo)** for fostering an environment of innovation and excellence in engineering and applied mathematics, which has been instrumental in the development of the computational and statistical frameworks underpinning this research.

Special thanks go to the **IME/USP (Institute of Mathematics and Statistics at the University of São Paulo)** for its rigorous academic training and research support, particularly in the areas of probability theory, game theory, and optimization techniques, which form the backbone of our quantitative model.

We would also like to acknowledge the **Rāmānujan Institute for the Development of Prodigious Young Mathematicians** for its commitment to nurturing mathematical talent and fostering groundbreaking research. The institute's emphasis on interdisciplinary collaboration has inspired the synthesis of diverse methodologies in this work.

Our gratitude extends to the technical teams at **EquiVerse** and **Holosystems**, whose expertise in data science, machine learning, and quantum computing has been invaluable. The technical corps at EquiVerse provided critical insights into equine genetics and performance analytics, while Holosystems contributed cutting-edge tools for large-scale data processing and simulation.

Finally, we thank our colleagues, collaborators, and the anonymous reviewers for their constructive feedback and support throughout the development of this research. This work would not have been possible without the collective effort of these institutions and individuals.

This research was supported by the **Rāmānujan Institute for the Development of Prodigious Young Mathematicians**.

Author Biographies

Marcos Eduardo Elias is a Brazilian mathematician, engineer, and entrepreneur with a distinguished career in finance and academia. He holds a PhD in Mathematics from St. Petersburg State University (Russia), where his research focused on stochastic processes and optimization techniques. Elias has held prominent positions in both academia and industry, including serving as a professor at institutions such as Ibmec, Insper, and FGV-SP. He co-founded several financial institutions, including GAS Investimentos (later Vinci Partners) and Empiricus Research. His work in quantitative finance and algorithmic trading has been widely recognized, and he is a frequent speaker at international conferences on financial modeling and risk management.

José Antonio Comegno Filho is an agricultural engineer graduated from FFALM - UENP (Foundation Colleges "Luiz Meneghel", State University of Northern Parana), MBA graduate at FGV (Fundacao Getulio Vargas). Family history of more with the horse feed trade. His research focuses on quantitative finance, market efficiency, and the integration of domain-specific data such as genetics, nutrition, and PED history into predictive models for horse racing betting.

Bibliography

- Benter, W. (1994). Computer-Based Horse Race Handicapping and Wagering Systems: A Report. *Proceedings of the Eleventh International Conference on Gambling & Risk Taking*.

- Forrest, D., & McHale, I. G. (2007). Public information and odds-on betting: The case of English football. *Applied Economics, 39*(10), 1193-1200.

- Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal, 35*(4), 917-926.

- Nash, J. (1950). Equilibrium Points in N-Person Games. *Proceedings of the National Academy of Sciences, 36*(1), 48-49.

- Ranogajec, Z. (2013). Large-scale betting and market efficiency: Insights from a professional syndicate. *Private Research paper*.

- Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature, 36*(4), 2021-2064.

- Taleb, N. N. (2012). *Antifragile: Things That Gain from Disorder*. Random House.
- Thaler, R. H., & Ziemba, W. T. (1988). Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives, 2*(2), 161-174.
- Thorp, E. O. (1969). Optimal Gambling Systems for Favorable Games. *Revue de l'Institut International de Statistique, 37*(3), 273-293.
- Woods, A. (1998). Quantitative Approaches to Race Betting. *Private syndicate research*.
- Ziemba, W. T., & Hausch, D. B. (1986). *Betting at the Racetrack*. Dr. Z Investments, Inc.
- Lo, A. W. (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. *Journal of Portfolio Management, 30*(5), 15-29.
- Beyer, A. (1993). *Picking Winners: A Horseplayer's Guide*. Houghton Mifflin Harcourt.
- Crist, S. (1986). *Exotic Betting: How to Make Multihorse, Multirace Bets That Win*. Daily Racing Form Press.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Silver, D., et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science, 362*(6419), 1140-1144.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Appendix: Python Code for Horse Racing Simulation

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from scipy.stats import levy, poisson, beta, cauchy, zipf

# Set random seed for reproducibility
```

```
np.random.seed(42)

# Define number of horses and iterations

num_horses = 10

num_iterations = 100000

# Define prior parameters for Beta distribution

alpha = np.ones(num_horses)  # Prior success count

beta_ = np.ones(num_horses)  # Prior failure count

# Simulate race outcomes

for _ in range(num_iterations):

    # Sample from Beta distribution to get win probabilities

    win_probs = np.random.beta(alpha, beta_)

    # Normalize probabilities to sum to 1

    win_probs /= win_probs.sum()

    # Simulate race outcome

    winner = np.random.choice(num_horses, p=win_probs)

    # Update Beta distribution parameters

    alpha[winner] += 1

    beta_[np.arange(num_horses) != winner] += 1

# Compute posterior win probabilities

posterior_probs = alpha / (alpha + beta_)

# Plot posterior win probabilities

plt.bar(range(num_horses), posterior_probs)

plt.xlabel('Horse')

plt.ylabel('Posterior Win Probability')

plt.title('Bayesian Win Probabilities')
```

```python
plt.show()
```

This code simulates a horse race using Bayesian updating and visualizes the posterior win probabilities. The model can be extended to incorporate additional factors such as track conditions, jockey skill, and domain-specific data.