

Estimating the Market Efficiency of Prediction Markets

Cyrus Chan CID: 02560148

Contents

1	Why this question	2
2	Background	2
3	Model	2
4	Data Observations	3
5	Deriving the MLE of θ	4
6	Analysis of findings	4
6.1	Favorite-Longshot Bias	4
6.2	Optimism bias	4
7	Conclusion	5
8	Critical Use of AI	5

1 Why this question

Prediction markets are platforms where participants can trade contracts based on the outcome of future events.[1] Prediction markets are said to be more accurate than expert opinions, as it magnifies the "wisdom of the crowd" to a larger scale. [1]

I chose this question as I was curious about the effectiveness and accuracy of analysing the prices of the contracts to predict the outcome of events that we are interested in.

2 Background

The prediction market platform we will retrieve data from in this report is Kalshi. Kalshi contracts are based on whether real-world events occur, with YES/NO options.

The sum of the price of the YES and NO contracts sum to \$ 1.00. If the outcome of the event aligns with the contract the user purchases, the user will receive \$ 1.00.[2]

The price of each contract is determined by employing a dynamic pricing model that directly correlates with the market's perceived probability of a specific event occurring.[3]

3 Model

The outcome of a contract is either it resolves to YES or NO, so we can model this as a Bernoulli distribution.

Let y_1, \dots, y_n be a sample of the outcomes related to Kalshi contracts. We can model these as realisations of a random variable Y_i for $i = 1, \dots, n$.

$$Y_i \sim \{\text{Bernoulli}(p_i) : p_i \in [0, 1]\}$$

We will take the market price of a contract at each time period as x_i cents. Since $x_i \in [0, 1]$, It can be said that if the market were perfectly efficient, $x_i = p_i$. Therefore, we will model each p_i as such:

$$p_i = x_i^\theta$$

$\theta < 1$	Price too low
$\theta = 1$	Perfectly efficient
$\theta > 1$	Price too high

As $\theta < 1 \Rightarrow x_i < p_i$ and $\theta > 1 \Rightarrow x_i > p_i$

4 Data Observations

The markets we will collect data from are respectively:

- 1.How many Grammy awards will Sabrina Carpenter win?[4]
- 2.How many Grammy awards will Lady Gaga win?[5]
- 3.How many Grammy awards will Kendrick Lamar win?[6]
- 4.How many Grammy awards will Bad Bunny win?[7]

These markets are all for the 2025 Grammy awards, and were chosen due to their similarity, so it can be assumed that the distribution their contract prices and event probability will be similar.

We will collect the data from 12 November 2025 to 25 January 2026, using the 'minute' frequency available on Kalshi. The actual date of the Grammy Awards was 1 February 2026. The date 25 January 2026 was chosen so that there would be less effects of high price fluctuations during the award shows. Each of these markets have different contracts, which respectively pay out based on the number of awards each performer wins, from exactly 0 to 9.

Performer	Actual number of Grammy Awards won
Sabrina Carpenter	0
Lady Gaga	2
Kendrick Lamar	5
Bad Bunny	3

x_i = Price of the contract at the given time

$$y_i = \begin{cases} 1 & \text{if the contract paid out} \\ 0 & \text{if the contract did not pay out} \end{cases}$$

Timestamp	Exactly 0	Exactly 1	Exactly 2	Exactly 3
2025-11-18 21:30:00	26.00	30.00	24.00	22.00
2025-11-19 01:15:00	26.00	36.00	24.00	22.00

Table 1: A snippet of the data being processed.[4] For example, this means that at 2025-11-19 21:30:00, the price for a contract predicting Exactly 0 awards won by Sabrina Carpenter was \$0.26

5 Deriving the MLE of θ

As our model of Y_i is a Bernoulli distribution, the model has probability function:

$$P(Y_i = 1|x_i; \theta) = p_i = x_i^\theta$$

The pmf of the distribution is:

$$f(y_i; \theta) = (x_i^\theta)^{y_i} (1 - x_i^\theta)^{1-y_i}$$

Hence the likelihood function of θ is:

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n (x_i^\theta)^{y_i} (1 - x_i^\theta)^{1-y_i}$$

With log-likelihood function of:

$$\log(L(\theta)) = \sum_{i=1}^n \log[(x_i^\theta)^{y_i} (1 - x_i^\theta)^{1-y_i}]$$

Using the scipy function `scipy.optimize.minimize`, we can find the minimum of $-\log(L(\theta))$ and hence find the θ with the maximum $L(\theta)$

Market	MLE of θ
Sabrina Carpenter	1.0804
Lady Gaga	1.2949
Kendrick Lamar	1.1229
Bad Bunny	1.1976

Table 2: Values for the minimum of the log-likelihood function calculated using python [8]

6 Analysis of findings

The data shown when calculating the MLE of θ seems to show that the price x_i of the contracts are consistently overpriced. Which means, the probability of the event actually happening was lower than what the prices would indicate on the Kalshi market. This indicates a systematic market inefficiency. There could be a few explanations for this:

6.1 Favorite-Longshot Bias

In prediction and betting markets, people overvalue longshots (outcomes with low probability), because the potential payout is much greater. In the context of horse-race beting, there is a systematic undervaluation of the odds of favourite winners, and an overvaluation of "dark horses". [9] In the context the data analysed here, it could be that traders are more inclined to purchase contracts with lower odds, leading to a positive bias to the prices.

6.2 Optimism bias

As this market specifically is based on the entertainment and music industry, unlike some other markets traded on Kalshi, such as those based on inflation or Federal Reserve interest rates, the markets analysed in this dataset could be traded by fans of the artists, which could lead to sentiment-driven trading, with a higher bias towards hoping the artists would win awards. This is especially prominent for the artists who won more awards: Lady Gaga, Kendrick Lamar and Bad Bunny.

These artists had a higher MLE of θ , as such, it implies that traders expected these artists to receive more awards. In contrast, there was a lower MLE value of θ for Sabrina Carpenter winning 0 awards.

7 Conclusion

For the markets analysed in this report, specifically related to the entertainment and music industry, there appears to be a significant inefficiency in the price of the contracts, as the price of the contracts appear to be systemically overpriced across the board.

The time span chosen is meant to be when there is less information regarding the award winners, due to cutting off the data analysed by around a week before the award shows. If the data was taken nearer to when the award show occurs, traders may have had more information to make better trades, and the price point would be closer to the true probability of the event occurring.

In conclusion, the market efficiency seems to be low in this case, and using the market price to predict the outcome of the events related to the contracts is likely unfeasible. However, the prices are better used as an estimate of the sentiment of the crowd.

8 Critical Use of AI

I used Google Gemini in order to debug and optimize the code I wrote, and to help format my LaTeX code. I also used it to help link my findings with the favorite-longshot bias.¹

References

- [1] Katelyn Peters. Prediction markets explained: Types, uses, and real-world examples. <https://www.investopedia.com/terms/p/prediction-market.asp>, 2026. Accessed: 2026-02-01.
- [2] Kalshi. Working with event contracts. <https://news.kalshi.com/p/what-are-event-contracts>, 2026. Accessed: 2026-02-01.
- [3] Kalshi. How are prices determined? <https://help.kalshi.com/markets/markets-101/how-are-prices-determined>, 2026. Accessed: 2026-02-02.
- [4] Kalshi. How many grammy awards will sabrina carpenter win? <https://kalshi.com/markets/kxgrammycountsab/award/kxgrammycountsab-26>, 2026. Accessed: 2026-02-02.
- [5] Kalshi. How many grammy awards will lady gaga win? <https://kalshi.com/markets/kxgrammycountlad/award/kxgrammycountlad-26>, 2026. Accessed: 2026-02-02.
- [6] Kalshi. How many grammy awards will kendrick lamar win? <https://kalshi.com/markets/kxgrammycountken/award/kxgrammycountken-26>, 2026. Accessed: 2026-02-02.
- [7] Kalshi. How many grammy awards will bad bunny win? <https://kalshi.com/markets/kxgrammycountbad/award/kxgrammycountbad-26>, 2026. Accessed: 2026-02-02.
- [8] Cyrus Chan. coursework-stats-modelling. <https://github.com/chc4036/coursework-stats-modelling>, 2026. Accessed: 2026-02-04.
- [9] R. M. Griffith. Odds adjustments by american horse-race bettors. *The American Journal of Psychology*, 62(2):290–294, 1949. ISSN 00029556. URL <http://www.jstor.org/stable/1418469>.