Do Sportsbooks Follow Bayesian Updating?

Nalin Khanna*

May 2023

Abstract

We study how sportsbooks update their odds over time. While recent literature has focused on price evolution in betting exchanges, the majority of sports betting volume occurs on sportsbooks. We analyze a custom dataset of odds from 12 sportsbooks. We test whether the stream of odds quotes follows the properties of Bayesian updating. For many sportsbooks, We find excess volatility in odds updates, inconsistent with a stream of Bayesian beliefs. Our conclusions extend previously documented excess volatility on betting exchanges to the realm of sportsbooks.

1 Introduction

A core assumption of many economic models is rationality. Agents are said to be rational if, in response to new information, they update their beliefs according to Bayes' Law and make decisions that maximize their expected utility under those beliefs. Beginning in the 1980s, the veracity of both of these assumptions began to come under attack. Informed by psychologists Kahneman and Tversky's *Thinking Fast and Slow*, behavioral economists began to question if economic agents are fully rational. One line of attack was that humans possess certain biases that could interfere with Bayesian updating. These biases include confirmation bias, base-rate neglect, and underreacting or overreacting to news in certain situations. Though these affronts to rationality have been documented in the stylized environments of the psychology lab, they are harder to prove in real-world settings. For example, one may suspect the stock market overreacts in response to news, and as such exhibits too much volatility. This claim, however, is quite hard to prove as we cannot have a researcher interrogate "the market" about its beliefs. In response to this challenge, behavioral economists, such as Robert Shiller, have constructed models of stock prices and shown that actual prices are too volatile with respect to what can be rationally explained by the model (Shiller 1981). However, this line of reasoning requires persuading critics that one's model truly captures reality. Instead, in a recent paper, Augenblick and Rabin (AR) explore model-free statistical tests of rational price movement [1].

*This project grew out of a Summer Research Project with fellow research assistants Gabriela Fuschini, Itamar Fayler, and Andrew Wang, and advised by Professor Philipp Strack. I would like to credit everyone on the team for their amazing contributions.

2 Literature Review

2.1 Augenblick and Rabin (2020)

Augenblick and Rabin's paper, which we will subsequently reference as AR, mixes both theory and application. In the theory part they develop a statistical test to check Bayesian updating, and in the empirical part they use this test on 3 datasets to see if they could have been generated from a Bayesian updater.

2.1.1 AR Statistical Test

The core of AR's test of Bayesian updating comes from martingale theory. Under a null hypothesis of Bayesian updating, a stream of beliefs $\pi_1, \pi_2, ..., \pi_T$ about whether an event will happen, will be a martingale, meaning $E[\pi_{t+1}|\pi_t, ..., \pi_1] = \pi_t$. In plain terms, part of updating rationally means one's expected future belief is reflected in their current belief. AR define two quantities over two arbitrary periods, t_1 and t_2, in a belief stream. The first is the movement, M, across periods which is the squared difference of the belief changes in the intervening periods between t_1 and t_2, $M_{t_1,t_2} = \sum_{i=t_1}^{t_2}(\pi_{i+1} - \pi_i)^2$, and the second is the uncertainty reduction, R, across periods, defined to be $R_{t_1,t_2} = \pi_{t_1}(1 - \pi_{t_1}) - \pi_{t_2}(1 - \pi_{t_2})$. AR's theoretical breakthrough is in relating movement and uncertainty reduction. They show that for any two periods, $E[M] = E[R]$ [1] That is, for a martingale, the expected movement across periods equals the expected uncertainty reduction[2].

Proposition 1 (Movement and Uncertainty Reduction Equivalence)**.**
For a stream of beliefs $\pi = \pi_1, \dots, \pi_T$ and arbitrary periods $1 \leq t_1, t_2 \leq T$:

$$E[M_{t_1,t_2}] = E[R_{t_1,t_2}].$$

The power of this result is that without making any assumptions about the underlying data-generating process inducing the beliefs, or requiring a reliance on a model for "correct" beliefs, the authors are still able to statistically reject the Bayesianness of a set of beliefs streams if they deviate from this $E[M] = E[R]$ property.

Still, $E[M] = E[R]$ is just an equation, how can we use this equivalence to generate a statistical test of Bayesian updating? The authors decide to focus on one-period belief changes, to maximize their sample size. For a given one period belief change, they look at the difference between M and R, termed "excess movement": $EM = M - R$. Under the null of Bayesian updating, each EM observation can be viewed as a draw from a mean-zero distribution, with some finite[3] variance σ^2. By the Central Limit Theorem, the average of n independent EM observations will converge in distribution toward $\overline{EM} \sim N(0, \frac{\sigma^2}{n})$, as $n \to \infty$. We can generate an unbiased estimate of σ^2 using the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(EM_i - \overline{EM})^2$.

[1] See appendix of AR for proof

[2] This equivalence is remarkable considering M and R have a profound asymmetry, R only depends on the displacement in belief between the start and final period while M depends on every step of the path taken from the start period to the final period.

[3] Given that beliefs are probabilities bounded between 0 and 1

Proposition 2 (Central Limit Theorem for Excess Movement)**.** *Given n independent excess movement statistics, $EM_1, \dots, EM_n$, and a null hypothesis that all EM_i have been generated from Bayesian belief updates,*

$$\overline{EM} \xrightarrow{d} N(0, \frac{s^2}{n})$$

Proposition 2 gives us a way to test $H_0 : \overline{EM} = 0, H_a : \overline{EM} \neq 0$ using a simple t-test. To do so, however, we must believe our dataset is big enough for the CLT Normal approximation. After doing numerical simulations, AR come up with the heuristic of requiring the sum of all movement in the data to be ≥ 3 for the CLT to be a strong approximation.

At this point, the reader may be asking themselves whether one-period belief changes within the same belief stream are really independent. For example, positive excess movement at the start of the match might predict positive excess movement later in the match, because it could be an indication of a volatile signal environment. The flaw in this reasoning is that We are operating under the null hypothesis that every belief in every stream is the result of a Bayesian update. As such, a positive EM observation at the start of the stream will not predict anything about the sign of the EM observations later in the stream since $E[EM] = 0$ for every one-period belief change, regardless of past EM values.

2.1.2 AR Data Analysis

With a statistical test in mind, We now move on to the data AR used. Authors performed their test on 3 datasets: a survey dataset of beliefs about world events, a dataset of live winning probabilities generated by a sports prediction algorithm, and another dataset of winning probabilities but this time from a sports betting exchange. For all three datasets, AR find excess movement to be statistically different from zero. For the survey data, they find movement to be 20% higher than uncertainty reduction. For the algorithmic data, they find movement to be 6.7% less than uncertainty reduction. And for the market data they find movement to be 4.6% more than uncertainty reduction. The results show strong disparities in bias direction and magnitude depending on the entity doing the updating.

3 Our Contribution

The rest of this paper will be about our work on top of AR. We extend AR's findings on betting markets to centralized sportsbooks. AR's sports betting results come from Betfair, which is an exchange where users interact directly with each other. Users trade contracts that pay out $1 if team A wins, and so the price of a contract is the probability the market is giving team A to win. This is an ideal setup from a research point of view as market beliefs are given directly by the price. The vast majority of betting volume, however, comes from centralized bookmakers, like Fanduel and DraftKings. On these platforms, users can only trade against the odds posted by the bookmaker. Our contribution is to map bookmaker's odds to market beliefs and then run AR's test on this dataset and compare the results to the exchange data.

4 Methodology

4.1 Odds Mapping

As we alluded to earlier, our dataset consists of odds quotes but AR's test is on beliefs, so we need to construct a mapping from odds to beliefs. Though odds are reported in various formats depending on the bookmaker, we can think of odds O on outcome A as the payoff of betting \$1 on outcome A, if A occurs. Then, the expected profit of betting \$1 on A is $z_A = p_A O_A - 1$, where p_A is the objective probability with which A occurs. We define A's implied probability $\pi(O_A)$ as likelihood of A that makes $z_A = 0$, thus making O_A "fair" odds. Solving for $z_A = 0$, We get $\pi_A = \frac{1}{O_A}$, π_A is known as the "implied odds" of O_A. We use π as our mapping from odds to beliefs with the interpretation that π_A is the market belief of the probability that outcome A will happen. By "market belief", we mean the belief of the marginal trader. In equilibrium, the marginal trader must be indifferent between buying and not buying outcome A at the posted odds O_A, and for them to be indifferent, their belief for the probability of outcome A must be π_A[4]. Thus, a stream of implied odds $[\pi_0(O_A), ..., \pi_T(O_A)]$, reflects a series of market belief updates, which we perform AR's Bayesian updating test upon.

There is, however, a drawback in this mapping, π, which is that if you take a set S of mutually exhaustive outcomes, e.g S = {Germany wins, England wins}, and sum over each outcome's implied probability at an arbitrary period t, you will get $\sum_{i \in S} \pi_t(O_i) > 1$, because bookmakers offer less than fair odds, to capture a profit. Thus, the implied odds do not form a probability distribution over S. In this way, it seems like using π as our mapping results in market beliefs that are inconsistent with each other, and cannot be viewed as probabilities. However, this inconsistency can be rationalized by realizing that there is no one market. The market for each outcome in S is segmented and has different marginal traders who hold different beliefs. The marginal trader m_G in the G ="Germany wins" market holds different beliefs than the marginal trader in the E ="England wins" market, because m_G holds belief $\pi(O_G)$ that Germany will win and therefore belief $1 - \pi(O_G)$ that England will win, but since $\pi(O_G) + \pi(O_E) > 1$, we know $1 - \pi(O_G) < \pi(O_E) = m_E$'s belief that England will win. As such, $\sum_{i \in S} \pi(O_i) > 1$ simply means each outcome market has a marginal trader with different beliefs. This can survive because the centralized sportsbook "buy-only" format prevents inter-outcome arbitrage opportunities.

5 Data

From August 2021 to December 2021, we used an API[5] to get odds data from 12 centralized bookmakers. The raw dataset contained 2,450,304 belief observations from 28,965 belief streams on 2,258 sporting events[6]. As we wanted to see how AR's results extended to sportsbooks, we followed their methodological decisions with our dataset, for comparability of results. As such, we only included streams from major American sports (Soccer (MLS), American Football (NFL and NCAAF), Baseball (MLB), and Basketball (NBA and NCAAB)). We only used beliefs from when

[4]Assuming traders are risk-neutral

[5]The API is called odds-api. Documentation is at https://the-odds-api.com/liveapi/guides/v4/.

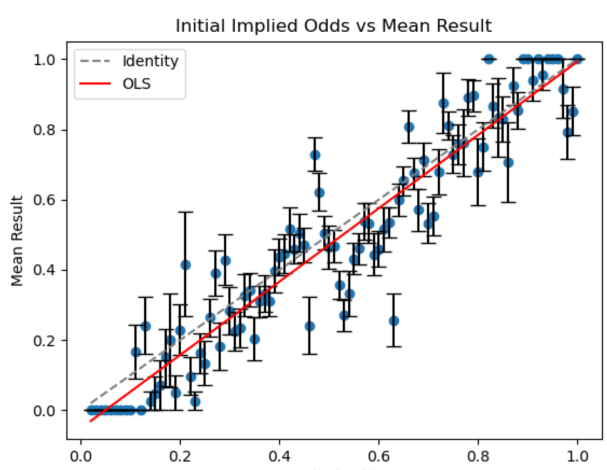
[6]Since different sportsbooks will quote odds for the same sporting event, we have many belief streams over the same events

the game was live. For each match, we only used the belief stream over which team would win and eschewed correlated streams such as whether a team would cover the moneyline. Sometimes, the API would report the same odds as the last query but with a more recent timestamp. Since AR's test is on changes in beliefs, we excluded those rows. This results in one observation every 2.5 minutes, on average. Thus, the interval between periods is roughly on the same timescale as the AR data, which had maximum one observation per minute. After all of this filtering, the dataset we did our statistical testing on had 356,973 belief observations from 5,220 belief streams on 672 sporting events. In comparison, AR's Betfair dataset contained 7 million observations on 300,000 sporting events. AR's data comes from 2006-2014, so our data is from a shorter but more recent period.

6 Descriptive Statistics

We look into some basic descriptive statistics of our data. One natural question to ask is, if the implied odds $\pi(O_A)$ reflect the marginal trader's belief of the likelihood that A will happen, does A actually happen $\pi\%$ of the time? To answer this, we regress the average occurrence of A on the initial belief for event A, i.e $p_a \sim \pi_0(O_A)$. In Figure 1, we see the regression line of best fit lies under the $p_a = \pi_0(O_A)$ line, meaning teams win less than their initial implied odds would suggest. This is unsurprising, as bookmakers give less than fair odds it makes sense that $p_A < \pi_0(O_A)$. What is interesting is that, according to the figure, the level of spread, $\pi_0(O_A) - p_A$ that the bookies charge changes depending on whether the team is a favorite or underdog. Bookies charge the most spread for longshot underdogs, and almost no spread for near-certain favorites. This is a well-documented phenomenon[7] and is known as longshot-favorite bias in the literature.

Figure 1: Test of Calibration for Implied Odds



Notes: Observations are grouped into 1% buckets so an initial belief of .265 and .274 are both part of the .27 bucket. Observations are presented with 95% CI error bars.

[7]For robustness We also run the same exercise on beliefs mid-game instead of at the start and find the same favorite-longshot bias (See Appendix).

Before performing the AR statistical test, it is also interesting to look at how the level of uncertainty reduction and movement changes over time in a match[8]. Augenblick, Lazarus, and Thaler (2023)[2] analyze this in the sportsbook setting, and we follow the methodology in their paper to generate our plots in Figure 2. We find that on average movement and uncertainty reduction are increasing over time for each sport, in agreement with ALT (2023). This finding makes sense as signals at the end of a match (i.e. a golden goal) can cause large belief changes (high movement) towards one of the outcomes (high uncertainty reduction).

7 Results

We present two test statistics. The first is the previously mentioned mean excess movement statistic: $EM = \frac{1}{n}\sum_i (m-r)^i_{t,t+1}$, where $(m-r)^i_{t,t+1}$ is the ith one-period excess movement observation. We also present what AR term "normalized excess movement" or $EM_{norm} = \frac{EM}{r_{t,t+1}} + 1$. This statistic is more interpretable, for example an $EM_{norm} = 1.4$ means movement was 40% higher than uncertainty reduction. The Z-score we report is still from testing $H_0 : EM = 0$. Our results are in figure 3.

Figure 3: Bayesian Updating Test Results by Sportsbook

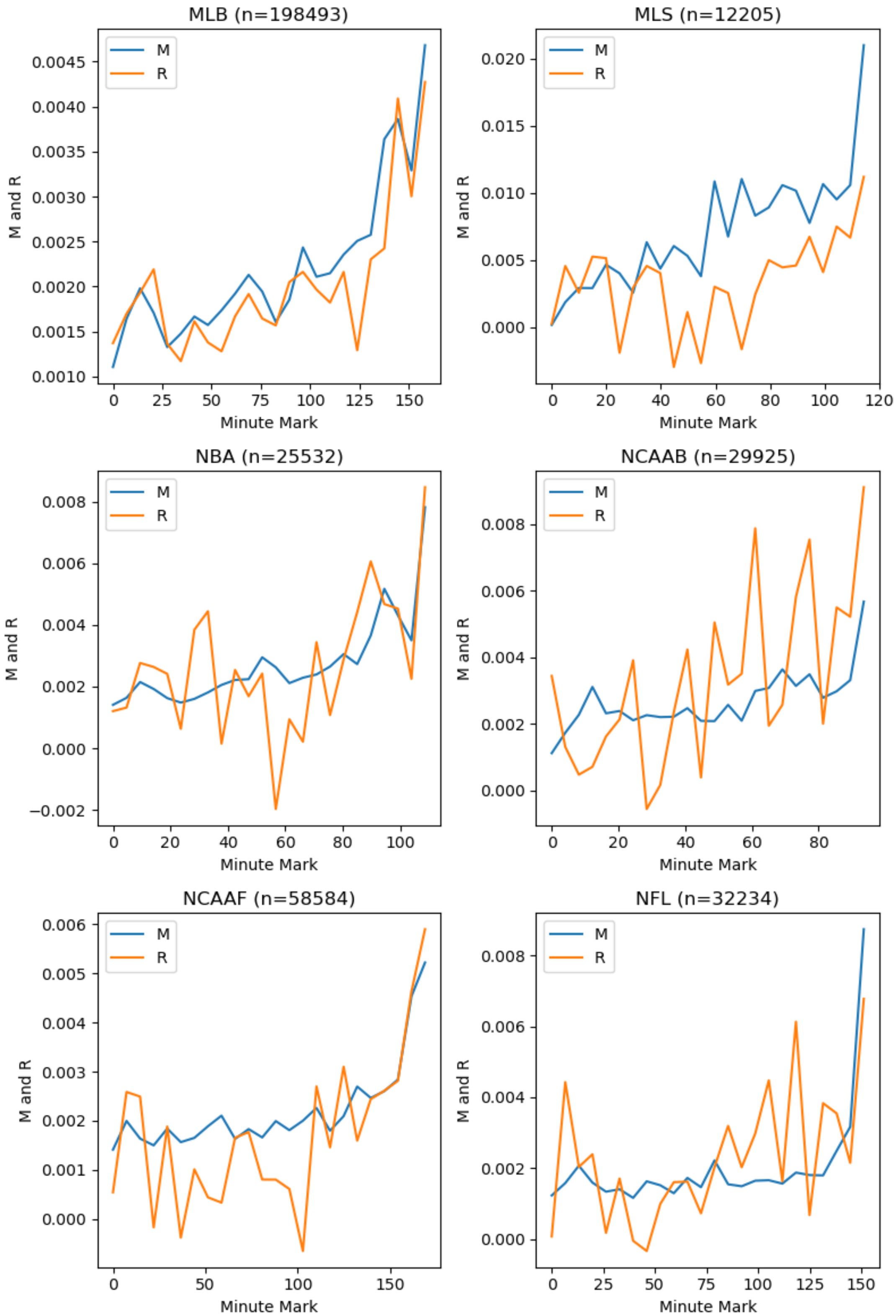
	Sportsbook	n	Total M	Average EM	EM_norm	Z score	p value
0	Barstool Sportsbook	35964	89.02850	0.00030	1.13635	2.47325	0.01339
1	BetMGM	23088	75.25900	0.00007	1.02336	0.44057	0.65953
2	Bovada	34793	98.28590	0.00036	1.14553	2.67373	0.00750
3	DraftKings	36380	94.79130	0.00020	1.08236	1.55166	0.12074
4	FOX Bet	21087	51.31080	0.00008	1.03440	0.55649	0.57787
5	FanDuel	39352	93.41430	0.00004	1.01747	0.33971	0.73408
6	GTbets	19137	49.28680	0.00042	1.19502	2.80126	0.00509
7	PointsBet (US)	19991	61.44630	-0.00000	0.99838	-0.02650	0.97886
8	SugarHouse	35334	87.00510	0.00028	1.13024	2.31516	0.02060
9	TwinSpires	33703	84.68960	0.00032	1.14547	2.49751	0.01251
10	Unibet	37268	92.56380	0.00028	1.12653	2.32315	0.02017
11	William Hill (US)	15617	57.94170	0.00014	1.03898	0.59576	0.55133

We present results at the sportsbook level, because different sportsbooks cover the same matches, and excess movement will be correlated for Bayesian observers sharing beliefs about the same stream. Instead, we rely on uncorrelated, independent belief changes to perform the AR test. The first thing to note is that total movement is much greater than 3 for all sportsbooks so the CLT approximation is valid for every row. Next, we see that all sportsbooks but one show positive excess movement in sample. Half (six) have excess movement positive enough to be significantly different from 0 at the 95% confidence level. We see normalized excess movement ranging from 1.00 to 1.20, meaning movement 0-20% greater than uncertainty reduction. AR's sample effect size falls within this range at with a normalized excess movement of 1.05.

In the six sportsbooks with significantly positive excess movement, these results are evidence that the marginal trader overreacts to signals, compared to a Bayesian observer. Positive excess movement means that trader beliefs are shifting too much in relation to what traders are actually

[8] We also plot average movement and uncertainty reduction by sportsbook in the appendix

Figure 2: Movement and Uncertainty Reduction over Time for Different Sports



learning. Normalized excess movement ranges from 13-20% in these sportsbooks, suggesting irrational updating is increasing volatility by about 15%.

8 Robustness

We check the robustness of our results along two axes: the first is our choice of odds mapping, and the second is our choice of outcome to track. We picked π as our odds mapping in order to interpret our odds data as the beliefs of the marginal trader in each outcome market. There are, however, other logical mappings from odds to probabilities and we want to make sure that our results are robust to the choice of mapping. As such, we rerun the main test using a different mapping:

$$\pi_t^*(O_A) = \frac{\pi_t(O_A)}{\sum_{i \in S} \pi_t(O_i)}$$

$\pi*$ deflates the implied odds π by $\sum_{i \in S} \pi_t(O_i)$, known as the "overround". The attractive property of $\pi*$ is that $\sum_{i \in S} \pi_t^*(O_i) = 1$, and thus π_t^* forms a probability distribution over the outcome set S. The economic motivation behind π_t^* is that it removes the distortion sportsbooks introduce in betting markets, by dividing out their profit margin. One downside of π_t^* is that probabilities under this mapping can no longer be interpreted as the equilibrium beliefs of a marginal trader, because those actually trading in the market do actually have to face the transaction cost of the overround. However, there is an alternative interpretation of $\pi_t^*(O_A)$, which is as the probability the sportsbook assigns to outcome A, which they then inflate by the overround to capture profit. But, as we saw in the calibration section, sportsbooks charge different levels of spread depending on whether a team is an underdog or favorite, so this interpretation is not fully accurate, but rather a good first-step approximation of bookmaker beliefs.

Figure 4 gives the results of AR's test on these approximate bookmaker beliefs:

Figure 4: Bayesian Updating Test Results on Deflated Beliefs by Sportsbook

	Sportsbook	n	Total M	Average EM	EM_norm	Z score	p value
0	Barstool Sportsbook	35964	80.10580	0.00028	1.14512	2.62453	0.00868
1	BetMGM	23088	67.40580	−0.00007	0.97640	−0.46266	0.64361
2	Bovada	34793	87.51820	0.00031	1.14187	2.61837	0.00884
3	DraftKings	36380	85.65140	0.00011	1.05048	0.97062	0.33174
4	FOX Bet	21087	45.75970	0.00006	1.02912	0.47791	0.63272
5	FanDuel	39352	82.33860	−0.00008	0.96230	−0.78897	0.43013
6	GTbets	19137	43.89620	0.00035	1.18277	2.64844	0.00809
7	PointsBet (US)	19991	55.89770	−0.00008	0.97249	−0.46157	0.64439
8	SugarHouse	35334	78.24100	0.00027	1.13808	2.44775	0.01438
9	TwinSpires	33703	76.19320	0.00029	1.14924	2.56962	0.01018
10	Unibet	37268	83.39800	0.00026	1.13300	2.43587	0.01486
11	William Hill (US)	15617	52.46170	−0.00002	0.99448	−0.08755	0.93024

As we can see, the results change but not by much. Excess movement decreases slightly for most (eight) sportsbooks. For three sportsbooks this is enough to push excess movement negative, but not statistically significantly so. The same six sportsbooks with significantly positive excess movement under π continue having significantly positive excess movement under $\pi*$. Thus, our conclusions are robust to a change in odds-to-belief mapping function.

For every sporting event, our API returns data on two outcomes: the odds that "Team 1" wins and the odds that "Team 2" wins (and the odds of a draw if its a soccer match). The API's choice of "Team 1" is simply the team whose name comes first alphabetically, so the designation is quite arbitrary. Up until now, we have done all of our analysis on the odds that Team 1 wins, ignoring the Team 2 odds stream since it is, of course, correlated with the Team 1 stream. However, if our results are capturing genuine positive excess movement by traders at certain sportsbooks, we would expect them to be robust to the choice of team to run AR's test on. So, for our second robustness check, we rerun the main test on the odds that Team 2 will win using the π mapping. The results are just below in figure 5:

Figure 5: Bayesian Updating Test Results on Team 2 by Sportsbook

	Sportsbook	n	Total M	Average EM	EM_norm	Z score	p value
0	Barstool Sportsbook	35964	87.36270	0.00025	1.11236	2.09026	0.03659
1	BetMGM	23088	71.51090	-0.00007	0.97871	-0.40871	0.68275
2	Bovada	34793	95.48830	0.00031	1.12608	2.34308	0.01913
3	DraftKings	36380	92.56280	0.00013	1.05377	1.03390	0.30118
4	FOX Bet	21087	51.52650	0.00003	1.01106	0.18288	0.85490
5	FanDuel	39352	89.87340	-0.00003	0.98631	-0.27060	0.78670
6	GTbets	19137	49.29860	0.00042	1.19455	2.76104	0.00576
7	PointsBet (US)	19991	57.71800	-0.00013	0.95730	-0.72190	0.47036
8	SugarHouse	35334	85.28320	0.00023	1.10576	1.92492	0.05424
9	TwinSpires	33703	82.91230	0.00025	1.11137	1.98142	0.04754
10	Unibet	37268	89.54260	0.00022	1.09865	1.84476	0.06507
11	William Hill (US)	15617	53.43940	-0.00005	0.98563	-0.22129	0.82487

The figure shows slight changes in the results. Normalized excess movement now ranges from 4.3% less than uncertainty reduction to 19.4% more. Two sportsbooks, Unibet and SugarHouse, go from significant positive excess movement to having their p-values slip just over .05, to .065 and .054, respectively. All other sportsbooks continue to either show significant positive excess movement or no significant excess movement. As such, we find our results are largely robust to the choice of outcome to track.

9 Conclusion

This paper analyzes whether sportsbooks update their odds following Bayes rule. It repurposes a recent procedure that researchers Augenblick and Rabin applied to betting exchange prices and applies it to sportsbook odds quotes. In half of the sportsbooks in our sample, we reject Bayesian updating due to significant positive excess movement of beliefs. This comports with Augenblick and Rabin's results of statistically significant positive excess movement on Betfair. Our normalized excess movement effect size for non-Bayesian sportsbooks ranges from 12.6-19.5%, which is even larger than the effect size of 4.6% that AR find for Betfair. Our results are robust to changes in odds mapping and outcome of interest.

Our paper is limited, however, by the assumptions made in order to go from odds to beliefs. We assume the existence of a risk-neutral marginal trader, whose beliefs stand in for the market's. In reality, many sports gamblers are risk-seeking over the small amounts they gamble over and some sophisticated professional gamblers are risk-averse. We find evidence of risk-seeking behavior in the data as our implied odds are miscalibrated downwards in Figure 1. A more sophisticated model of risk preferences could elicit a more accurate mapping from odds to trader's subjective beliefs.

Future research could also look into whether arbitrage opportunities exist on sportsbooks with positive excess movement. Positive excess movement captures a tendency of beliefs to revert to $\frac{1}{2}$ which suggests buying outcomes with implied odds less than $\frac{1}{2}$, i.e. underdogs . Of course, this strategy would have to generate more expected value than is being lost to the spread (which is the worst for underdogs), in order to be profitable.

In conclusion, we find that the excess volatility documented on Betfair from 2006-2014 by Augenblick and Rabin (2020), also exists on Barstool, Bovada, Unibet, and other top sportsbooks in 2021.

10 Appendix

Figure 6: Robustness Test of Calibration for Implied Odds

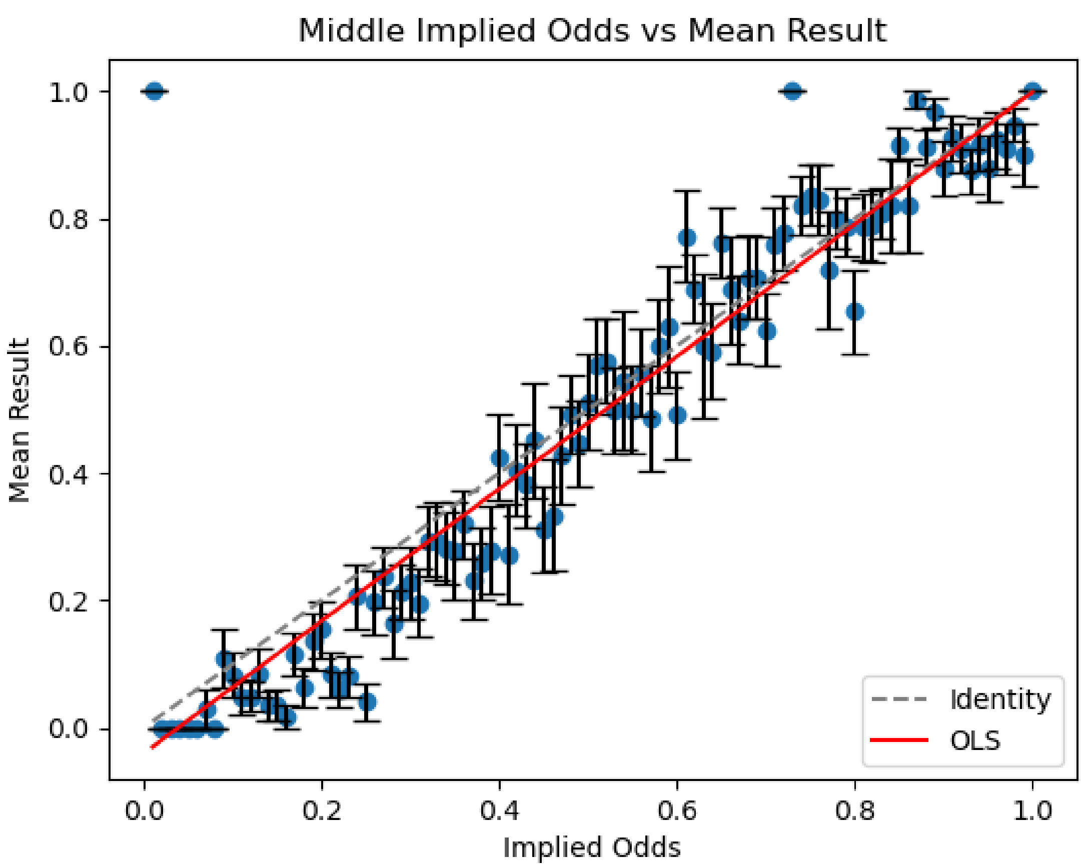
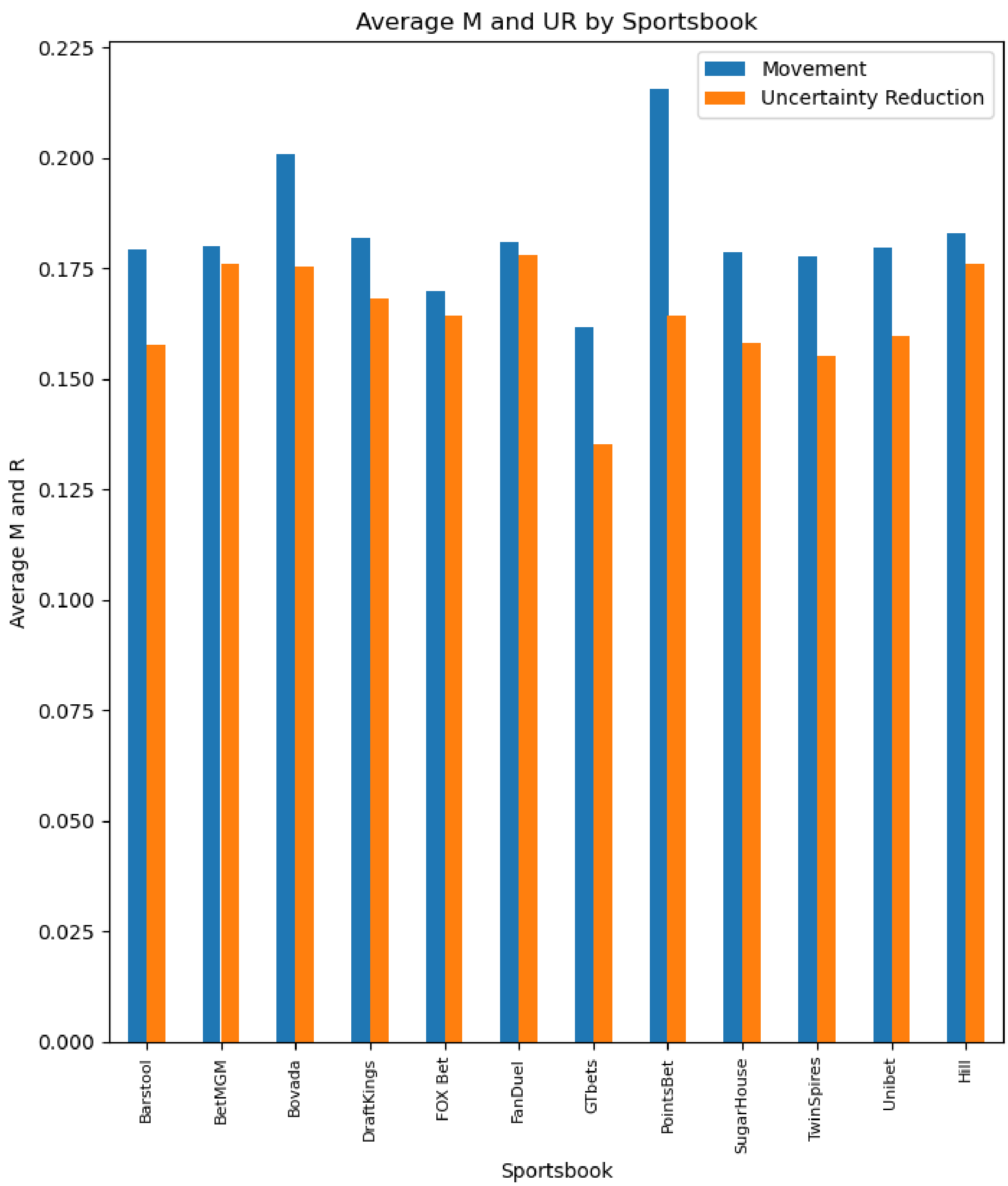


Figure 7: Average Movement and Uncertainty Reduction by Sportsbook



References

[1] Ned Augenblick and Matthew Rabin. "Belief Movement, Uncertainty Reduction, and Rational Updating". In: *The Quarterly Journal of Economics* (2021), pp. 933–985.

[2] Ned Augenblick Eben Lazarus and Michael Thaler. *Overinference from Weak Signals and Underinference from Strong Signals*. March 2023.