Vince Laird

SPORTS BETTING LINE MOVEMENT
DATA SCIENCE - CAREER TRACK

Introduction

Sports betting lines move throughout the day, and it would be useful to be able to predict line movement. This will help people betting on sports make an informed decision on the optimal time of when to place their wager.

Solving the problem of bad beats and betting at the wrong time

ESPN posted an article on their website detailing "The worst 'bad beats' of the 2016-17 NFL season". The phrase 'bad beat' is often used to describe an action late in a contest that results in an individual going instantly from winning a bet to losing. Focusing on the game featured second in the article:

> Philadelphia Eagles at Dallas Cowboys
>
> Closing line: Cowboys -5 (in other words, had a bettor placed a bet on the Cowboys at the closing "number" of -5, and by "closing" we mean right before the game begins, the Cowboys would have to win the game by 6 points; had a bettor placed a bet on the Eagles, the Eagles would have to lose the game by no more than 4 points; a game ending in a 5 point advantage for the Cowboys would result in a "push", which voids all payouts but allows bettors to recoup their original bet amount).
>
> The Eagles took a 23-13 lead early in the fourth quarter, and those who took the underdog were feeling great. After all, if you factor in the spread, the Eagles were up 15. ... The Cowboys scored the next 10 points to send the game to overtime. Dallas won the toss and marched right down the field, capping off the drive with a Jason Witten 5-yard touchdown catch to win by six.

Final score: Cowboys 29, Eagles 23 in OT (Cowboys win by 6 and "cover", making a winner of anyone who bet -5 on the Cowboys and a loser of anyone who bet +5 on the Eagles.)

Could this 'bad beat' have been avoiding? Perhaps - though there are many different betting options in Las Vegas to place a bet on this contest, we'll examine the line movement at South Point:

		Money Line		Spread		Total	
Date	Time	Fav	Dog	Fav	Dog	Over	Under
10/24	1:34pm			DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110
10/26	12:29pm	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110
10/27	10:25am	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110
10/28	10:55am	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.0 -110	43.0 -110
10/29	1:28pm	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.0 -110	43.0 -110
10/29	3:35pm	DAL-220	PHI+190	DAL-4.5 -110	PHI+4.5 -110	43.0 -110	43.0 -110
10/29	3:47pm	DAL-220	PHI+190	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/29	3:52pm	DAL-230	PHI+195	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	11:12am	DAL-230	PHI+195	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	11:38am	DAL-230	PHI+195	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	11:39am	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	12:12pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	43.5 -110	43.5 -110
10/30	1:04pm	DAL XX	PHI XX	DALXX XX	PHIXX XX	XX XX	XX XX
10/30	1:09pm	DAL XX	PHI XX	DALXX XX	PHIXX XX	XX XX	XX XX
10/30	4:25pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	43.5 -110	43.5 -110
10/30	7:07pm	DAL-250	PHI+210	DAL-5.5 -110	PHI+5.5 -110	44.0 -110	44.0 -110
10/30	7:29pm	DAL-250	PHI+210	DAL-5.5 -110	PHI+5.5 -110	44.5 -110	44.5 -110
10/30	7:36pm	DAL-250	PHI+210	DAL-6 -110	PHI+6 -110	45.0 -110	45.0 -110
10/30	7:43pm	DAL-250	PHI+210	DAL-6 -110	PHI+6 -110	45.0 -110	45.0 -110
10/30	7:46pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	45.0 -110	45.0 -110
10/30	7:56pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	45.0 -110	45.0 -110
10/30	8:06pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	45.0 -110	45.0 -110
10/30	8:21pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	44.5 -110	44.5 -110
10/30	8:24pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	44.5 -110	44.5 -110
10/30	9:58pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	44.5 -110	44.5 -110
10/30	10:01pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	44.5 -110	44.5 -110

The columns to note above are: Date, Time, and "Spread" (both "Fav" and "Dog"). Let's review the first row of data, in the screenshot on the following page:

		Money Line		Spread		Total	
		SOUTHPOINT LINE MOVEMENTS					
Date	Time	Fav	Dog	Fav	Dog	Over	Under
10/24	1:34pm			DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110

This line "opened" at 1:34 PM on 10/24, which was a Monday. The contest was scheduled to begin at 8:30 PM on 10/30, a Sunday evening matchup. Data on when the contest is scheduled to begin is not displayed, however this information can be found on several websites which show the start times and dates for all NFL football contests.

The opening line for this contest was: Dallas Cowboys -4.5, the Cowboys were expected to win by 4 ½ points. The -110 that follows DAL-4.5 represents what is often referred to as the "juice", "vig" (short for vigorish), or "odds". This value represents the fraction of the wager taken from the bettor as sort of a holding cost: in this example, a winning wager of $110 will yield a profit of $100. This number occurs after the Dallas line and the Philadelphia line in the screenshot above (it usually is -110 for spreads, but not always). This house advantage is easier to understand when you consider what would occur if it didn't exist, i.e.:

Two people bet at South Point on the Philadelphia vs. Dallas contest. There is no "vig" so each person places their wager and will receive "even money" if they win - they each bet $100 and each stands to win $100. Person A bets on Philadelphia, Person B bets on Dallas. These two bets are the only two wagers placed on the contest; the game is played, Dallas wins in convincing fashion and Person B wins their bet. Person B collects $200 (the amount they won plus the amount of their original wager), while Person A collects $0.

In the above scenario, the casino keeps nothing. South Point took in $200 in wagers from Person A and Person B combined, but 100% of those wagers were paid out to Person B. This is not the way Vegas operates! And this is why the vig exists.

Now imagine the same scenario with -110 odds on each bet. Person A now must bet $110 to win $100, and does so, again on Philadelphia. Person B also bets $110 to win $100, on Dallas. Again, Dallas wins easily. Now, Person B collects $210 (the amount they won plus the amount of their original wager), while Person A collects $0. The total amount wagered on this contest was $220 though: $110 bet by two people. Total payouts were $210. The $10 difference is the vig or juice, and that is the fraction that the casino keeps.

To answer the question of what problem this project aims to solve, it's this one: if you want to bet on Philadelphia between when the line opens at 1:34 PM on 10/24 and when the contest begins at 8:30 PM on 10/30, when do you place your bet?

This contest was only a bad beat for bettors who did not bet on Philadelphia during the time that the line was +6, which did occur, from 7:36 PM to 7:46 PM on 10/30, less than an hour before kickoff. Granted, this is not a huge window of time, but if you were considering making a bet on Philadelphia on 10/24 and there was data suggesting that you'd receive more advantageous odds if you waited, maybe you should wait.

		Money Line		Spread		Total	
Date	Time	Fav	Dog	Fav	Dog	Over	Under
10/24	1:34pm			DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110
10/26	12:29pm	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110
10/27	10:25am	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.5 -110	43.5 -110
10/28	10:55am	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.0 -110	43.0 -110
10/29	1:28pm	DAL-200	PHI+175	DAL-4.5 -110	PHI+4.5 -110	43.0 -110	43.0 -110
10/29	3:35pm	DAL-220	PHI+190	DAL-4.5 -110	PHI+4.5 -110	43.0 -110	43.0 -110
10/29	3:47pm	DAL-220	PHI+190	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/29	3:52pm	DAL-230	PHI+195	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	11:12am	DAL-230	PHI+195	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	11:38am	DAL-230	PHI+195	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	11:39am	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	43.0 -110	43.0 -110
10/30	12:12pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	43.5 -110	43.5 -110
10/30	1:04pm	DAL XX	PHI XX	DALXX XX	PHIXX XX	XX XX	XX XX
10/30	1:09pm	DAL XX	PHI XX	DALXX XX	PHIXX XX	XX XX	XX XX
10/30	4:25pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	43.5 -110	43.5 -110
10/30	7:07pm	DAL-250	PHI+210	DAL-5.5 -110	PHI+5.5 -110	44.0 -110	44.0 -110
10/30	7:29pm	DAL-250	PHI+210	DAL-5.5 -110	PHI+5.5 -110	44.5 -110	44.5 -110
10/30	7:36pm	DAL-250	PHI+210	DAL-6 -110	PHI+6 -110	45.0 -110	45.0 -110
10/30	7:43pm	DAL-250	PHI+210	DAL-6 -110	PHI+6 -110	45.0 -110	45.0 -110
10/30	7:46pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	45.0 -110	45.0 -110
10/30	7:56pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	45.0 -110	45.0 -110
10/30	8:06pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	45.0 -110	45.0 -110
10/30	8:21pm	DAL-240	PHI+200	DAL-5.5 -110	PHI+5.5 -110	44.5 -110	44.5 -110
10/30	8:24pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	44.5 -110	44.5 -110
10/30	9:58pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	44.5 -110	44.5 -110
10/30	10:01pm	DAL-240	PHI+200	DAL-5 -110	PHI+5 -110	44.5 -110	44.5 -110

SOUTHPOINT LINE MOVEMENTS

Why would anyone care about optimal betting time

For the same reason that people often try to get a better price/better deal on items they are purchasing, my belief is that if data were presented in a logical, objective fashion that could be proven to help sports bettors make more informed decisions about when to wager, some bettors would be interested.

I've seen it mentioned that less than 10% of gamblers profit from sports betting - no idea if that's true but it sounds about right. And yet, people continue to do it.

My opinion is that people who are interested in betting on sports probably realize that it's a losing venture. They do it for fun. If people had information that could potentially give them an edge in this losing venture, maybe to tilt the scales just a little bit in their favor, they'd consider it - if the data was presented in a interesting, easy-to-understand manner.

What data can be used for this, and how will it be acquired

There is historical line movement data available on the internet; the table in this document showing line movement on the contest between Philadelphia and Dallas was found on vegasinsider.com by manipulating their url (their urls appear to follow a certain pattern).

For current line movement, there is a service that I currently subscribe to called jsonodds.com - this site allows for API calls to be made and the site returns current odds on various sports. I am planning on creating a script that will make calls to this API throughout the day and store the data returned in a database.

How will this problem be solved

Time series analysis. If there is a "market" for sports betting lines, I have to wonder how efficient it is; there may be virtually no pattern in line movements at all, I haven't analyzed the data or looked at enough lines to note volatility or other significant patterns. There are "betting adages", such as "bet favorites early and bet dogs late" but I have no idea how accurate these are.

There are certainly outside factors that could affect line movement that should be considered in some way: if a key player is ruled out due to injury, or simply getting a day off, that news would not become publically available at any consistent time or in any consistent way. This news could be made available by the player themselves, such as on their Twitter feed, or the team may make an official announcement on their website.

Another factor would be weather, for games played outdoors. A classic example was last year's Notre Dame vs. North Carolina State college football game which was played during Hurricane Matthew. The predicted total points scored by both teams combined opened at 65 three days prior to kickoff. The total had moved down to 56.5 by the day of the game which was likely due to current weather reports and the belief that both teams would have difficulty playing football in such horrible weather conditions. This ended up being the case as the game ended with a score of 10-3.

Deliverables

My best case scenario would be to build a website which would have predictive tables and charts for all contests across all sports, updating in real time.

Realistically? Probably not going to happen like that, but I would like to have some sort of application that could at least be queried to show somewhat updated information on sports betting line movement predictions on games scheduled in the future, even if it's only for one sport.

Sources

The ESPN article about bad beats:

http://www.espn.com/chalk/story/_/id/18636571/nfl-super-bowl-li-six-worst-gambling-bad-beats-2016-17-nfl-season

The historical odds on the Dallas/Philadelphia game:

http://www.vegasinsider.com/nfl/odds/las-vegas/line-movement/eagles-@-cowboys.cfm/date/10-30-16/time/2030

Sports bettors only winning 10% of their bets? (It's probably even worse than this.)

http://www.biasharainsight.com/2016/06/you-cant-make-money-in-sports-betting/

The historical odds on the Notre Dame/NC State game:

http://www.vegasinsider.com/college-football/odds/las-vegas/line-movement/notre-dame-@-north-carolina-state.cfm/date/10-8-16/time/1200

What happens when you try to play a football game during a hurricane:

http://www.usatoday.com/story/sports/ncaaf/2016/10/08/hurricane-matthew-nc-state-notre-dame-ncaa-football/91785756/

Sports betting adage, analyzed:

https://www.sportsinsights.com/blog/should-you-really-bet-nfl-favorites-early-and-underdogs-late/