Offense vs. Pitching: An Econometric Analysis of Winning in Major League Baseball

Adrian Montez

ECON 5283-999: Data Visualization with Python

October 26, 2025

Abstract

Major League Baseball teams face pressure every year to leverage financial and player development resources efficiently to maximize on-field success. A central question arises each season: should a team prioritize improving offensive performance or strengthening its pitching staff to increase its winning percentage? This study seeks to evaluate which factor is more strongly associated with overall team success in the modern era (2015-2024), team offensive performance (as measured by Weighted Runs Created Plus, wRC+), or team pitching performance (as measured by Fielding Independent Pitching Minus, FIP-), while controlling for team payroll. There will be an additional robustness check using alternative offensive (On-Base-Plus-Slugging Plus, OPS+) and defensive (Earned Runs Allowed Minus, ERA-). The analysis finds that pitching performance is more strongly associated to team winning percentage than offensive performance over the 2015-2024 period, as confirmed by a formal test rejecting the hypothesis of equal impact. These results will provide insights into how teams may evaluate trade-offs and demonstrate the use of econometric methods to address performance-outcome questions in sports analytics.

Offense vs. Pitching: An Econometric Analysis of Winning in Major League Baseball

1. Introduction

Every year, each Major League Baseball team finishes its season with one question in mind: How can we improve? The answer is seldom clear, as there are countless options for enhancing the team, and choosing whether to prioritize offense or pitching is complex. Baseball teams, especially those with smaller budgets, face many trade-offs because of this. For example, investing in a strong offensive outfielder could mean missing out on a pitcher whose production demands a large payout. These problems require economic solutions, as the choice to sign one player over another could be the deciding factor in whether a team has a winning record worthy of making the playoffs. This paper aims to provide insight into this key issue by examining whether productive offense or pitching is more strongly related to winning in the modern, analytical era of Major League Baseball (2015 to 2024).

To formally evaluate whether offense or pitching contributes more to winning, this paper frames the question as a test of comparative marginal effects. Specifically, the null hypothesis assumes that improvements in offense and pitching influence winning percentage equally in magnitude (but in opposite directions due to scaling conventions), whereas the alternative hypothesis assumes that the predictor variables influence winning percentage unequally in magnitude, as shown below.

$$H_0: \beta_{wRC+} = -\beta_{FIP-}$$
$$H_1: \beta_{wRC+} \neq -\beta_{FIP-}$$

As seen in the hypothesis statements, offensive production will be measured through the Team wRC+ (Weighted Runs Created Plus) statistic, while pitching production will be measured with Team FIP- (Fielding Independent Pitching Minus). It should also be noted that this analysis will be controlling for payroll to ensure that teams with high payrolls (such as the Dodgers, Mets, Yankees, etc.) do not inflate the effects of offense and pitching on wins, and isolating performance differences as measured by standardized metrics. Additionally, there will be robustness checks with alternative metrics: On-Base-Plus-Slugging Plus (OPS+) and Earned Runs Allowed Minus (ERA-).

Many studies examine the effects of offense and pitching on team wins. Ault (1984) found that pitching and defense were slightly more important than offense in determining team wins. However, this study was conducted in an era of baseball when analytics were not utilized and embraced to the extent they are today, making the landscape quite different. Ault also looked at different statistics, such as batting average and earned run average (ERA). Ehrlich and Potter (2021) conducted a similar study with more recent data (2010-2017), but looked into the effect of offensive, pitching, and defensive production on team revenue using the Wins Above Replacement (WAR) metric.

2. Data

The dataset for this study was compiled using Fangraphs for team-level offensive, pitching, and win statistics over the ten-year period. Team payroll data was obtained from Spotrac, and additional offensive metrics were sourced from Sport Reference's Stathead subscription service. After importing the data into Google Colab, unnecessary columns were removed, and the datasets were merged using consistent "Season" and "Team" identifiers, including renaming a "Year" column in the payroll dataset to ensure compatibility. The 2020 season was excluded due to its shortened 60-game schedule. Winning Percentage (WPct) was then calculated as wins divided by total games played, and payroll was additionally transformed using the natural log to avoid any scaling-related issues. The final dataset was organized and checked for missing values, duplicate observations, and incomplete team-season records, with no issues detected.

Table 1: Descriptive Statistics

	Season	W	L	WPct	wRC+	FIP-	OPS+	ERA-	Payroll	log_Payroll
count	270.00000	270.00000	270.00000	270.00000	270.00000	270.00000	270.00000	270.00000	270.00000	270.00000
mean	2,019.44444	80.97778	80.97778	0.49999	97.72222	100.05556	97.89630	100.04074	144,273,876.64074	18.70913
std	3.02816	13.04323	13.02412	0.08047	9.62304	8.43963	8.91286	11.03805	56,702,754.54537	0.40306
min	2,015.00000	41.00000	51.00000	0.25309	75.00000	75.00000	76.00000	71.00000	42,421,870.00000	17.56317
25%	2,017.00000	72.00000	72.00000	0.44444	91.00000	95.00000	92.00000	92.00000	99,595,787.50000	18.41663
50%	2,019.00000	81.50000	80.00000	0.50464	97.00000	100.00000	97.50000	99.50000	140,318,618.00000	18.75942
75%	2,022.00000	90.75000	90.00000	0.55760	104.00000	106.00000	104.00000	107.00000	179,283,155.25000	19.00448
max	2,024.00000	111.00000	121.00000	0.68519	126.00000	124.00000	126.00000	133.00000	346,277,812.00000	19.66275

After cleaning the data, summary statistics were generated and are shown in Table 1. The observation counts confirm the correct number of data points in each column (30 teams' data tracked from 2015 to 2024, excluding 2020). The mean for our independent variables reveals interesting patterns early on, showing wRC+ has a mean of 97.722, slightly below the normalized league average of 100, indicating marginally lower offensive production across the sample period, which could mean a slight right skewness in the wRC+ data. In contrast, FIP- (another normalized statistic), has a mean of 100.056, showing that pitching performance is centered closely around the league average and more evenly distributed across teams. The standard deviations of wRC+ (9.623) and FIP- (8.44) show that, across teams and seasons, most of the values fall within 10 points of the average. Also, since the standard deviation is less than ten percent of the mean for both statistics, it suggests that the data points are consistent and tightly clustered around the mean. Payroll, however, has a very high standard deviation. Well over ten percent of its mean, it shows that there is a high level of variance in team payroll. This is further demonstrated in the quartiles for payroll, as the maximum is over $346 million while the minimum is only $42 million. This was expected, as a few teams (such as the Dodgers, Yankees, Mets, etc.) spend extremely high amounts compared to most teams. Lastly, the winning percentage's mean being roughly 0.5 also shows the structural balance of wins and losses across the MLB.

Table 2: Description of Variables

Variable	Data Type	Quantitative Type	Purpose
WPct	Float	Continuous	Dependent Variable
wRC+	Integer	Continuous	Main offensive variable
FIP-	Integer	Continuous	Main pitching variable
OPS+	Integer	Continuous	Robustness check
ERA-	Integer	Continuous	Robustness check
Payroll	Integer	Continuous	Control Variable
ln(Payroll)	Float	Continuous	Control Variable

It is also necessary to establish the rationale for selecting these variables and how they measure team-level offensive and pitching performance. Table 2 shows each of my variables to be used in the regression, along with their data type, quantitative type, and their purpose.

Winning percentage is the clear choice for the dependent variable, as it is a direct measurement of a team's success in a given season. Although it's derived from count data (the number of wins divided by the total number of games in a season), it behaves as a continuous variable appropriate for regression.

wRC+ is a comprehensive statistic that measures offensive production by weighing each offensive outcome (e.g., single, home run, strikeout) according to its value in creating runs. Additionally, it is normalized around a league average of 100 and adjusts for important external factors. FIP- is a statistic that strips away the effect of a team's defense and luck on balls put into play and focuses solely on the three "true outcomes" for a pitcher (strikeouts, walks, home runs). It is also normalized and adjusts for important external factors. The reason for choosing these as the primary variables is their ability to isolate a team's performance to "pure" offense and pitching. They are two of the most comprehensive statistics, and the above adjustments allow for cross-team and cross-season comparisons, making it possible to compare a team in 2015 to a different team in 2024. These skill-isolating statistics are more closely associated with underlying performance ability, and tend to be more stable across time than "outcome-based" metrics, such as OPS+ and ERA-, which will be the alternative metrics to determine the robustness of the results.

To explain these alternative metrics: OPS+ is an offensive statistic that adds On-Base Percentage and Slugging Percentage relative to the league average, while ERA- measures the amount of earned runs a team allows per nine innings relative to the league average. Because these statistics are less precise and can be influenced by external factors, such as a team's defense, they can produce noisier results. At the same time, these statistics still reflect what *actually* happened, while wRC+ and FIP- are more representative of what is *expected* to happen. If my results are to hold under both sets of metrics, there will be a strong case of robustness and evidence that the findings won't be a product of using only theoretical data. This distinction between predictive and realized performance is central to my analysis, as it allows me to test whether the side of the game that is *expected* to lead to more wins (based on underlying skill) *actually* does when compared to the side that did lead to more wins in the observed results.

Lastly, I chose to use team payroll as a control variable because it is a confounding variable. Logically, it would make sense that teams with higher payrolls can afford better hitters and pitchers, making it related to both the independent and dependent variables. Additionally, controlling for payroll can further isolate the relative explanatory power of offensive and pitching performance. Controlling for this essentially acts as a measure and representation of the available resources to a given team. Additionally, since team payroll is highly right-skewed and varies widely across organizations, payroll is transformed using the natural logarithm. This transformation will improve stability in the model and allows the coefficient to be interpreted as the relationship between proportional changes in payroll and winning percentage.

3. Methodology

To quantify the marginal associations between team offense, pitching, and winning percentage in the MLB, I estimate a baseline ordinary least squares (OLS) regression in which team winning percentage is modeled as a function of wRC+, FIP-, and team payroll. The model is represented by the following:

$$WPct_{i,t} = \beta_0 + \beta_1\, wRC+_{i,t} + \beta_2\, FIP\text{-}_{i,t} + \beta_3\, \ln(Payroll_{i,t}) + \varepsilon_{i,t}$$

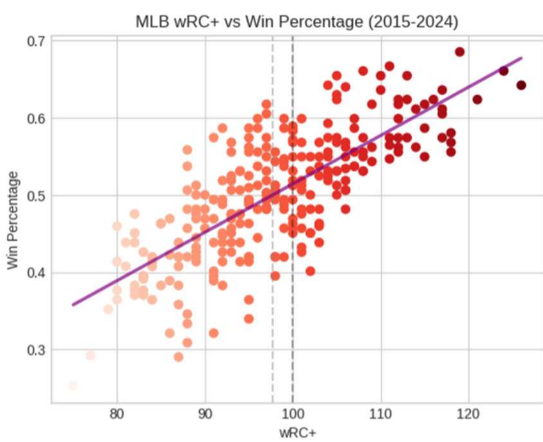
This regression is meant to quantify the marginal relationship between wins and the two skill areas at a team level (offense and pitching), while holding payroll constant, where i = team and t = year. Payroll is entered in logarithmic form to reflect proportional changes in spending and diminish the effects of any scale-related instability. It should be expected that the coefficient β_1 is positive, as stronger offensive production should increase winning percentage, while β_2 should be negative, since a lower FIP- should also increase a team's win rate. I expect strong relationships between winning percentage and both skill areas, so it should be necessary to test if the difference between the results is statistically significant. There is also an error term, $\varepsilon_{i,t}$, which captures unobserved factors not included in the model. I found the OLS regression to be appropriate as there is a continuous dependent variable (win percentage) and a large cross-sectional dataset. Although an OLS regression only identifies correlations, the robustness check with OPS+ and ERA- in place of wRC+ and FIP- assesses whether the observed relationships persist under alternative model specifications (OPS+ and ERA- will essentially replace wRC+ and FIP- in a second regression). It should also be noted that because the analysis relies on observational data, it cannot establish causal relationships between the performance metrics and wins.

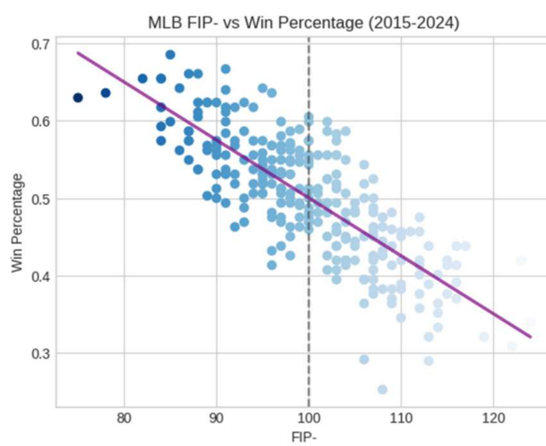
In addition to the regression, examining the data's visual trends can prove to be helpful. Scatter plots were prepared with a line of best fit to see each independent variable's linear relationship to the dependent variable and the strength of the relationship relative to each other. A second set of scatter plots to confirm that payroll is a confounding variable was also created.

4. Results

Getting into the visual trends in the data, Figure 1 shows a generally strong and positive linear relationship between wRC+ and winning percentage, meaning that better-hitting teams typically have a higher winning percentage. The light gray line indicates the mean, while the darker line indicates the normalized average. Meanwhile, Figure 2 shows that FIP- has a strong and negative linear relationship, meaning that better-pitching teams usually have a higher winning percentage (this is because FIP- being lower means better pitching). Comparing the strength of these relationships to each other based on these graphs may be too difficult, as both skill variables have roughly the same slope and strength in their relationship to winning percentage.

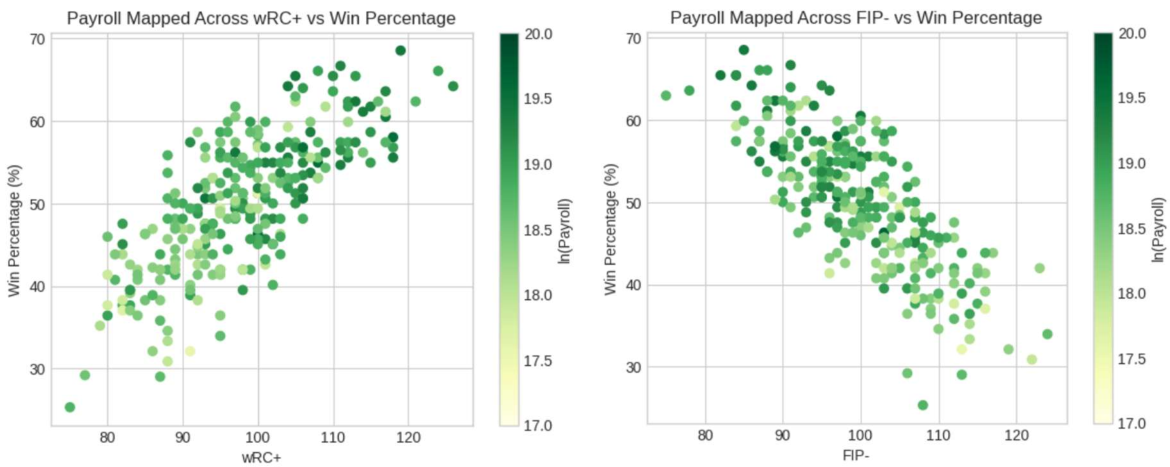
Figure 1 (left) & Figure 2 (right)





Figures 3 and 4 show how taking the natural log of Payroll visually compresses the values. These scatterplots show each team's payroll level, with a green dot indicating a higher payroll and a yellow dot indicating a lower payroll, while plotting the wRC+ (and FIP-) for each team according to their winning percentage. These graphs show that proportional differences in spending are relatively small compared to the variation in on-field performance.

Figure 3 (left) & Figure 4 (right)



As for the regression results, Table 3 presents the results of the main regression. All predictor variables are statistically significant, with p-values effectively zero ($p < 0.05$), indicating that the model as a whole explains a meaningful portion of the variation in winning percentage, and the coefficient estimates can also be safely interpreted. wRC+ has a coefficient of 0.004, meaning that for each one-point increase in team wRC+, the win percentage increases by 0.004 on average. Conversely, FIP- has a coefficient of -0.0052, indicating that as pitching performance worsens by one point (FIP- increases by one point), win percentage decreases by 0.0052 on average. On the surface, the results show that pitching shows a larger estimated marginal association with team success compared to offense, providing the first direct answer to the study’s primary question.

Table 3: Main Regression Results

OLS Regression Results

Dep. Variable:	WPct	R-squared:	0.782
Model:	OLS	Adj. R-squared:	0.779
Method:	Least Squares	F-statistic:	317.5
Date:	Fri, 12 Dec 2025	Prob (F-statistic):	1.43e-87
Time:	23:19:52	Log-Likelihood:	503.20
No. Observations:	270	AIC:	-998.4
Df Residuals:	266	BIC:	-984.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6314	0.051	12.350	0.000	0.531	0.732
wRC+	0.0040	0.000	13.407	0.000	0.003	0.005
FIP-	-0.0052	0.000	-16.215	0.000	-0.006	-0.005
Payroll	8.746e-07	4.6e-05	0.019	0.985	-8.98e-05	9.15e-05

Omnibus:	1.732	Durbin-Watson:	2.019
Prob(Omnibus):	0.421	Jarque-Bera (JB):	1.420
Skew:	0.137	Prob(JB):	0.492
Kurtosis:	3.226	Cond. No.	4.56e+03

Payroll, however, has a p-value of 0.654, showing no statistically significant effect on win percentage once offensive and pitching performance are accounted for. This suggests that wRC+ and FIP- capture team performance extremely well, and payroll's influence on success operates primarily through its impact on these skill metrics rather than directly. This reinforces that the comparative effects of wRC+ and FIP- are not confounded by payroll, reinforcing confidence in the estimated relationships.

The model fits the data well, with an R-squared of 0.782 and an adjusted R-squared of 0.779, indicating that the independent variables explain approximately 78% of the variation in win percentage. The narrow 95% confidence intervals further reinforce the precision in the estimates, as wRC+ ranges from 0.003 to 0.005 and FIP- ranges from -0.006 to -0.005. Neither interval includes zero, highlighting the statistical significance of these predictors.

While the condition number is high (7,780), indicating the possibility of strong multicollinearity, the Variance Inflation Factor (VIF) for each variable is below 2. This indicates that scaling issues may persist even after payroll is measured in the millions, yet the regression results remain interpretable.

A heteroscedasticity test was also conducted, and the results (p-value 0.067 > 0.05) show no strong evidence of heteroscedasticity, further supporting the reliability of the OLS estimates. Ensuring homoscedasticity further supports the reliability of the estimated difference between offensive and pitching effects and rules out the possibility that it is driven by inconsistent variance.

Finally, an F-test of coefficient equality in magnitude was performed to examine whether the magnitudes of the offense and pitching coefficients significantly differ, based on the hypotheses outlined in the introduction. The test yields a p-value of 0.017, allowing us to reject the null hypothesis and conclude that pitching performance exhibits a significantly stronger marginal association with winning percentage than offense.

The second regression (Table 4) using the alternative metrics (OPS+ and ERA-) yields highly similar results and passes all diagnostic checks, providing stronger empirical support. The coefficient estimates are 0.0036 for OPS+ and -0.005 for ERA-, reflecting a similar difference in magnitude. The corresponding F-test yields a p-value of 0.000161, further supporting the conclusion that pitching has a larger marginal impact on team success than offense. Notably, the R-squared value was higher for this model (0.874), meaning that the second model explains a larger share of the variation in team winning percentage. This is likely because win percentage and these outcome-based metrics both incorporate realized performance. Because the alternative metrics lead to the same conclusion, the finding that pitching has a stronger marginal association on team success proves robust across multiple model specifications.

Table 4: Regression Results Using Alternative Metrics

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   WPct   R-squared:                       0.874
Model:                            OLS   Adj. R-squared:                  0.872
Method:                 Least Squares   F-statistic:                     613.5
Date:                Sat, 13 Dec 2025   Prob (F-statistic):          3.65e-119
Time:                        00:07:03   Log-Likelihood:                 577.09
No. Observations:                 270   AIC:                            -1146.
Df Residuals:                     266   BIC:                            -1132.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.6391      0.034     18.913      0.000       0.573       0.706
OPS+           0.0037      0.000     15.606      0.000       0.003       0.004
ERA-          -0.0050      0.000    -27.356      0.000      -0.005      -0.005
Payroll    -5.463e-06   3.48e-05     -0.157      0.875    -7.4e-05     6.3e-05
==============================================================================
Omnibus:                        1.881   Durbin-Watson:                   1.944
Prob(Omnibus):                  0.390   Jarque-Bera (JB):                1.644
Skew:                          -0.086   Prob(JB):                        0.439
Kurtosis:                       3.341   Cond. No.                     3.96e+03
==============================================================================
```

5. Conclusion

This paper examined whether offensive production or pitching performance is more strongly associated with team success in Major League Baseball. Using team-level data from 2015 to 2024 and an OLS regression framework, the results show that both offense and pitching are statistically significant predictors of winning percentage. However, pitching performance, measured by FIP-, has a significantly larger marginal impact on winning than offensive performance measured by wRC+, even after controlling for team payroll. A formal F-test rejects the hypothesis that the two effects are equal in magnitude, providing strong evidence that pitching contributes more heavily to team success in the modern, analytical era of baseball.

These findings are robust to alternative predictor variables. When outcome-based metrics (OPS+ and ERA-) replace the primary skill-isolating measures, the results remain consistent, and the relative importance of pitching persists. Controlling for payroll further clarifies the relationship by accounting for differences in team resources. Once on-field performance is included, payroll does not have a statistically significant direct effect on winning percentage, suggesting that financial resources primarily influence outcome through performance rather than independently.

Several limitations should be acknowledged. The analysis relies on observational data and cannot establish causal relationships between the performance metrics and wins. The study also uses team-level aggregates, which may mask important variation within an individual team, such as player-specific contributions or mid-season roster changes. Additionally, defensive performance is not explicitly modeled, and future research could incorporate defensive team metrics. Lastly, this study does not include any team or season fixed effects that would control for any unobserved, unchanging differences across organizations (ex. market size) or any league-wide shocks (ex. rule changes). Future research could extend this analysis by incorporating more granular data and fixed-effects specifications.

From a decision-making standpoint, the results suggest that pitching performance is more strongly associated with success in baseball compared to offensive performance, even after accounting for payroll. These findings may help inform internal evaluations and performance diagnostics for teams, while illustrating the value of econometric methods in addressing strategic performance-outcome questions in sports analysis.

References

Ault (1984), “Pitching, Defense Just Slightly More Important to Team Wins Than Offense”. *1984 Baseball Research Journal*, https://sabr.org/journal/article/pitching-defense-just-slightly-more-important-to-team-wins-than-offense/

Ehrlich JA, Potter JM (2021), "Is offense worth more than defense and pitching? Marginal revenue product and revenue sharing in major league baseball". *Managerial Finance*, Vol. 47 No. 6 pp. 760–778, doi: https://doi.org/10.1108/MF-04-2020-0210

Major League Baseball (2015-2024), “Batting Team Data Dashboard”. *Published by FanGraphs*, https://www.fangraphs.com/leaders/major-league?team=0%2Cts&type=8&startdate=&enddate=&month=0&ind=1&season1=2015&season=2024&sortcol=0&sortdir=default&pagenum=1&pageitems=2000000000. Accessed on 9/14/2025

Major League Baseball (2015-2024), “Pitching Team Data Dashboard”. *Published by FanGraphs*, https://www.fangraphs.com/leaders/major-league?pos=all&stats=pit&lg=all&qual=y&type=8&month=0&ind=1&startdate=&enddate=&season1=2015&season=2024&pageitems=2000000000&team=0%2Cts. Accessed on 9/14/2025

Major League Baseball (2015-2024), “Pitching Team Data Advanced Dashboard”. *Published by FanGraphs*, https://www.fangraphs.com/leaders/major-league?team=0%2Cts&type=1&startdate=&enddate=&ind=1&season1=2015&season=2024&pageitems=2000000000&month=0&pos=all&stats=pit. Accessed on 9/14/2025

Major League Baseball (2015-2024), “Team Batting Season Stats”. *Published by Sports Reference*, https://stathead.com/baseball/team-batting-season-finder.cgi?request=1&order_by=year_id&year_min=2015&year_max=2024. Accessed on 9/15/2025

SportsDataIO (2015-2024), “MLB Team Salary Payroll Tracker”. *Published by Spotrac*, https://www.spotrac.com/mlb/payroll/_/year/2024. Accessed on 9/15/2025