THE VIRTUE OF TRANSPARENCY IN SPORTS ANALYTICS: AN APPLICATION TO NFL QUARTERBACKS

THIS VERSION FEBRUARY 16, 2026

Megan Czasonis is a Founding Partner of Cambridge Sports Analytics in Cambridge, MA.
245 Main Street, Cambridge MA, 02142
mczasonis@csanalytics.io

Miles Kee is a data scientist at Cambridge Sports Analytics in Cambridge, MA.
245 Main Street, Cambridge MA, 02142
mkee@csanalytics.io

Mark Kritzman is a Founding Partner of Cambridge Sports Analytics in Cambridge, MA, and a senior lecturer at the MIT Sloan School of Management in Cambridge, MA.
245 Main Street, Cambridge MA, 02142
mkritzman@csanalytics.io

David Turkington is a Founding Partner of Cambridge Sports Analytics in Cambridge, MA.
245 Main Street, Cambridge MA, 02142
dturkington@csanalytics.io

Abstract

A model-based approach to prediction, such as a regression model or a machine learning model, estimates parameters from a sample of observations and applies those parameters uniformly across all prediction tasks. These approaches fail to consider the unique circumstances of individual prediction tasks, and they conceal the influence of observations and predictive variables on individual predictions. Additionally, model-based approaches to prediction offer no guidance about the reliability of individual predictions. We describe a novel model-free prediction system called relevance-based prediction (RBP) that overcomes these limitations, and we show how it enables prediction-specific transparency that is beyond the reach of model-based approaches. RBP reveals the specific reliability of each prediction before the prediction is made, the importance of each prior player to each prediction, the contribution of each predictive variable to each prediction's value, and the contribution of each predictive variable to each prediction's reliability. We illustrate this new prediction system by applying it to predict passing yards per game for NFL quarterbacks. The prediction-specific information given by RBP stands in contrast to R-squared, t-statistics, and beta which only give information about average effects, as we illustrate with specific player examples.

THE VIRTUE OF TRANSPARENCY IN SPORTS ANALYTICS: AN APPLICATION TO NFL QUARTERBACKS

We describe a rigorous, intuitive, and model-free prediction system called relevance-based prediction (RBP) and illustrate its virtues by predicting passing yards per game for NFL quarterbacks. RBP forms transparent predictions that adapt to the unique circumstances of each prediction task, such as the prediction of a specific performance outcome for one chosen player. RBP forms a prediction as a precise weighted average of previously observed outcomes that occurred for other relevant players and situations using a rigorously defined statistic called relevance. Unlike models such as linear regression analysis or machine learning models, which work by estimating model parameters and then applying those parameters to form predictions, RBP operates by evaluating patterns in the relationship between outcomes and predictive variables given the specific circumstances of each prediction task.

Like machine learning models, RBP captures nonlinear effects and conditionality, but unlike machine learning models, RBP's model-free approach to prediction gives remarkable visibility into the formation of each prediction:

- It reveals the reliability of each prediction before the prediction is made.
- It reveals precisely how each player informs each prediction.
- It shows how each predictive variable contributes to each prediction's reliability and its value.

The prediction-specific information given by RBP stands in stark contrast to the summary statistics given by a model. For example, linear regression analysis provides one R-

squared, one set of t-statistics, and one set of beta coefficients. None of these statistics distinguish among the circumstances of different prediction tasks. By contrast, RBP's assessment of reliability, a variable's contribution to a prediction's reliability, and a variable's contribution to a prediction's value, reflect the player-specific context of each prediction task.

We begin by describing the three key components of the RBP system: relevance, fit, and grid prediction.[1] We then describe how RBP measures the influence of the predictive variables on a prediction's reliability and its value. Next, we apply RBP to predict passing yards per game for NFL quarterbacks. We describe our experiment setup, and we present results which give evidence of the transparency and efficacy of RBP. We conclude with a summary.

Relevance-Based Prediction

As described comprehensively by Czasonis, Kritzman, and Turkington (2022a, 2022b, 2023, and 2024), RBP is a model-free prediction system that forms a prediction as a weighted average of previously observed outcomes. The weights are specific to the single prediction task under consideration. The weights are determined in a precise mathematical fashion, and they offer extreme transparency into how the prediction is formed, which promotes intuition and understanding by both technical and nontechnical stakeholders. At a high-level, the process works as follows. First, for a chosen set of predictive variables and a chosen fraction of observations to include, RBP computes a rigorously defined relevance score for each observation, selects the fraction of observations that are most relevant, and assigns a weight to each observation accordingly. Second, RBP computes a measure of reliability, called fit, which reflects the strength of patterns in the data revealed by these specific weights for the current prediction task. Third, RBP applies the same approach to a large prediction grid of different combinations of predictive variables and different subsamples of observations to include and blends these calibrations together according to their relative reliability to form composite grid prediction weights and an aggregate grid prediction.

Relevance

Relevance provides a statistical measure of the importance of a previous player to the prediction for a current player given a chosen set of predictive variables. It is composed of two components, similarity and informativeness, which are both computed as Mahalanobis distances (Mahalanobis 1936) owing to principles of information theory and the importance of properly accounting for the variances and correlations of all the predictive variables. In equations 1 through 4, x_i is a row vector of the values of the predictive variables for a previous player, x_t is a row vector of the values of the predictive variables for the current player, $\bar{x}$ is a vector of the average values of the predictive variables for all previous players in the sample, Ω^{-1} is the inverse covariance matrix of the values of the predictive variables for all previous players, and $'$ denotes matrix transpose. Notice that similarity is the negative of a distance and that it must be scaled by $1/2$ to match the units of informativeness.[2] Intuitively, observations that are more different than average contain more information because their patterns are less likely to reflect noise. All else being equal, previous players who are like the current player but different from the average of all previous players are more relevant to a prediction than those who are not.

$$r_{it} = sim(x_i, x_t) + \frac{1}{2}\left(info(x_i, \bar{x}) + info(x_t, \bar{x})\right) \qquad (1)$$

$$sim(x_i, x_t) = -\frac{1}{2}(x_i - x_t)\Omega^{-1}(x_i - x_t)' \quad (2)$$

$$info(x_i, \bar{x}) = (x_i - \bar{x})\Omega^{-1}(x_i - \bar{x})' \quad (3)$$

$$info(x_t, \bar{x}) = (x_t - \bar{x})\Omega^{-1}(x_t - \bar{x})' \quad (4)$$

This definition of relevance is not arbitrary. In addition to its motivation by information theory, whereby the Mahalanobis distance reflects the information contained in a point on a multivariate normal distribution, relevance also converges to linear regression analysis as a special case. If we define weights as in equation 5, which admits the relevance-weighted average of every previous player outcome in the observed sample, and we form a prediction using equation 6, the result is equivalent to the prediction that results from linear regression analysis.[3]

$$w_{it,linear} = \frac{1}{N} + \frac{1}{N-1} r_{it} \quad (5)$$

$$\hat{y}_t = \sum_{i=1}^{N} w_{it} y_i \quad (6)$$

Owing to this equivalence, the theoretical justification given by Gauss for linear regression analysis applies as well to RBP. In most cases, however, we can produce a more reliable prediction by taking a relevance-weighted average of a subsample of relevant players, especially if the relationship between the predictive variables and the outcomes is not static and symmetric. RBP censors the influence of previous players who are less relevant than a chosen threshold, which leads to the following definition of prediction weights.

$$w_{it,partial} = \frac{1}{N} + \frac{\lambda^2}{n-1}(\delta(r_{it}) r_{it} - \varphi \bar{r}_{sub}) \quad (7)$$

$$\delta(r_{it}) = \begin{cases} 1 & if\ r_{it} \geq r^* \\ 0 & if\ r_{it} < r^* \end{cases} \quad (8)$$

$$\lambda^2 = \frac{\sigma^2_{r,full}}{\sigma^2_{r,partial}} = \frac{\frac{1}{N-1}\sum_i r_{it}^2}{\frac{1}{n-1}\sum_i \delta(r_{it}) r_{it}^2} \quad (9)$$

In equations 7 through 9, $n = \sum_{i=1}^{N}(r_{it})$ is the number of players who are fully retained, $\varphi = n/N$ is the fraction of players in the retained sample, and $\bar{r}_{sub} = \frac{1}{n}\sum_{i=1}^{N} \delta(r_{it}) r_{it}$ is the average relevance value of the players in the retained sample. Weights formed in this manner will always sum to one[4] and depend crucially on the prediction circumstances x_t.

<u>Fit</u>

RBP's measure of fit is an essential component of the prediction system and represents a distinct concept that is not observable in classical statistics or machine learning. Fit quantifies the prevalence of useful patterns in a dataset for the purposes of a specific prediction task. It reveals how much confidence we should have in a specific task, separately from the confidence we have in the overall prediction system. Thus, fit also provides a principled way to evaluate the relative merits of alternative calibrations for a prediction. Most importantly, fit is computed prior to making a prediction based only on the current prediction weights and corresponding previously observed player outcomes. There are two equivalent ways to compute fit: (1) the product of standardized weights and outcomes averaged across every pair of previous player observations in the data sample to gauge alignment, or (2) the squared correlation between weights and previous player outcomes.

$$fit_t = \frac{1}{(N-1)^2}\sum_i \sum_j z_{w_{it}} z_{w_{jt}} z_{y_i} z_{y_j} \quad (10)$$

$$fit_t = \rho(w_t, y)^2 \quad (11)$$

Like relevance, fit is not arbitrary. The informativeness-weighted average fit across all prediction tasks in the observed sample equals R-squared for the special case of linear regression we noted earlier.[5]

$$R^2 = \frac{1}{T-1}\sum_t info(x_t)fit_t \quad (12)$$

This convergence reveals that prediction-specific fit is the fundamental building block of R-squared, where R-squared is a model average that includes predictions of varying quality. To compute fit and obtain this prediction-specific insight into reliability, we must know the weight of each player in a prediction. These weights are inherent to RBP, but they are not available in model-based prediction algorithms which rely exclusively on calibrated parameters rather than weighted players to form predictions.

This notion of prediction-specific fit warrants particular emphasis. Because it offers advance guidance about a specific prediction's reliability, it enables teams to view with greater caution predictions that are foreseen to be unreliable.

<u>Grid Prediction</u>

We have thus far shown how to form a prediction as a relevance-weighted average of player outcomes. And we have shown how we can use fit to measure the reliability of a specific prediction task. But we have been assuming a fixed choice of predictive variables and fraction of observations to include. We now introduce grid prediction to consider every subset of predictive variables and multiple choices for r^*, the threshold which determines how many observations to censor. But first, we must describe an enhanced version of fit called adjusted fit.

To the extent there is asymmetry in the data that informs a prediction, it is prudent to trust the more relevant sample on principle. Prediction calibrations that censor irrelevant data to address asymmetry should get credit for doing so. We measure asymmetry based on the squared difference in correlation for prediction weights based on the retained sample $(+)$ and those that would have been formed based on the complementary censored sample $(-)$.

$$asymmetry_t = \frac{1}{2}\left(\rho\left(w_t^{(+)}, y\right) - \rho\left(w_t^{(-)}, y\right)\right)^2 \tag{13}$$

We add asymmetry to fit and multiply this sum by K, the number of predictive variables, to compute adjusted fit. Multiplication by the number of predictive variables allows us to compare predictions based on different numbers of predictive variables.

$$adjusted\ fit_t = F_t = K(fit_t + asymmetry_t) \tag{14}$$

Adjusted fit addresses the fundamental tradeoff that is inherent to prediction. Censoring irrelevant variables and observations may lead to clearer patterns, but as the sample becomes smaller and more focused it introduces additional noise and raises the risk of overfitting to spurious patterns. All else being equal, adjusted fit penalizes calibrations that rely on fewer predictive variables because small subsets of variables are more likely to identify spurious relationships. And all else being equal, adjusted fit penalizes calibrations that rely on smaller, more focused sets of player observations because the zero or near-zero weights of many players has a dampening effect on fit. Censored calibrations are appealing to the extent they reveal compelling patterns that more than compensate for these baseline penalties.

We now return to the question of how to form a prediction given uncertainty in the calibration of r^* and variable selection, which are codependent choices. To address this issue, we compute a composite prediction as a reliability-weighted average of the predictions from all possible calibrations. Equations 15 through 17 define reliability weights, ψ_θ, in terms of the adjusted fit for each calibration, θ, and show how the overall grid prediction as well as the weight assigned to each prior player observation is computed as a reliability-weighted average.

$$\psi_\theta = \frac{F_\theta}{\sum_{\tilde{\theta}} F_{\tilde{\theta}}} \quad (15)$$

$$\hat{y}_{t,grid} = \sum_\theta \psi_\theta \hat{y}_{t,\theta} \quad (16)$$

$$w_{it,grid} = \sum_\theta \psi_\theta w_{it,\theta} \quad (17)$$

Exhibit 1 gives a visual representation of grid prediction based on hypothetical values. The columns represent different combinations of predictive variables, and the rows represent different subsamples of previous players as determined by different relevance thresholds. Each cell represents a calibration θ; that is, a unique combination of predictive variables and previous players. In practice, we would consider all 63 combinations of six variables, but for illustrative purposes we show only seven columns in exhibit 1. The first values shown in the cells are the calibration-specific predictions $\hat{y}_t$ for a given prediction task t. The second values are the weights ψ_θ (shown in units of percentage points) that we apply to the calibration-specific predictions to form the composite prediction. Remember that the values in this grid represent one specific prediction task and would be recomputed for another prediction task. This illustration gives a composite prediction of 16.30 (15.7 x 1.72% + 15.7 x 1.15% + 10.1 x 0.24% + . . . + 9.3 x 0.04%).

Exhibit 1: Grid Prediction – Illustrative Example

Variable Combinations

Player Censoring Thresholds	$X_1 X_2 X_3 X_4 X_5 X_6$		$X_1 X_2 X_3 X_4$		$X_1 X_3 X_4$		$X_2 X_5 X_6$		$X_3 X_6$		X_2		X_6	
0.0	15.7	1.72	15.7	1.15	10.1	0.24	15.3	1.37	10.9	0.54	15.3	0.47	7.4	0.06
0.1	16.4	2.02	16.7	1.39	10.4	0.23	15.4	1.88	12.5	0.73	15.5	0.50	7.7	0.04
0.2	17.5	2.20	17.4	1.43	10.3	0.18	15.4	1.91	12.6	0.64	15.5	0.44	7.9	0.05
0.3	17.8	2.17	17.7	1.43	10.5	0.20	15.5	2.24	12.6	0.62	15.5	0.42	7.9	0.05
0.4	18.2	2.29	18.0	1.50	10.6	0.22	15.4	2.18	12.7	0.65	15.5	0.41	8.1	0.07
0.5	18.6	2.50	18.2	1.58	10.7	0.25	14.3	2.50	12.8	0.70	15.3	0.41	8.1	0.06
0.6	18.7	2.47	18.4	1.61	10.7	0.23	15.4	1.21	13.1	0.73	15.4	0.42	8.8	0.10
0.7	19.0	2.47	18.8	1.63	10.7	0.19	15.4	2.20	12.9	0.62	15.4	0.41	8.7	0.07
0.8	19.4	2.32	19.1	1.50	11.5	0.20	15.3	2.04	13.7	0.57	15.5	0.37	8.6	0.04
0.9	19.5	1.26	18.8	0.81	12.9	0.22	15.5	1.73	14.0	0.32	15.3	0.25	9.3	0.04

Composite prediction: 16.30

The prediction grid also yields a comprehensive measure of how each variable contributes to a prediction's reliability. Specifically, the Relevance-Based Importance (RBI) of variable k is given by the average adjusted fit, $F_{t\theta}$, for grid cells that include variable k ($\Delta_k(\theta) = 1$) minus the average adjusted fit for the remaining cells ($\Delta_k(\theta) = 0$), as shown by equation 18.

$$RBI_{tk} = \sum_{\theta} \left(\frac{\Delta_k(\theta) F_{t\theta}}{\sum_{\tilde{\theta}} \Delta_k(\tilde{\theta})} - \frac{(1-\Delta_k(\theta)) F_{t\theta}}{\sum_{\tilde{\theta}} \left(1-\Delta_k(\tilde{\theta})\right)} \right) \tag{18}$$

RBI has several advantages over alternative measures of variable importance. Linear regression analysis relies on t-statistics and their corresponding p-values, which only measure a variable's marginal importance. RBI, by contrast, captures a variable's total importance. RBI also captures conditional relationships which t-statistics fail to address. And unlike the Shapley value, which is the accepted standard for assessing variable importance in machine learning models, RBI accounts for the reliability of individual predictions.

We can use the same approach to measure how each variable contributes to the value of a prediction. This measure is called contribution to prediction (CTP) and it is given by the average prediction from the cells including variable k minus the average prediction from the remaining cells, as shown by equation 19.[6]

$$CTP_{tk} = \sum_{\theta} \left(\frac{\Delta_k(\theta)\hat{y}_{t\theta}}{\sum_{\tilde{\theta}} \Delta_k(\tilde{\theta})} - \frac{\left(1-\Delta_k(\theta)\right)\hat{y}_{t\theta}}{\sum_{\tilde{\theta}}\left(1-\Delta_k(\tilde{\theta})\right)} \right) \tag{19}$$

A final note on grid prediction. For some prediction tasks, it may be preferable to select the subsample of players and predictive variables based on similarity rather than relevance. We need not worry whether we should use similarity or relevance to identify the optimal combination of players and variables. We simply include both censoring rules as candidates in the grid. However, even when we censor players based on similarity, we should still form the predictions as a relevance-weighted average of the retained players.

Experiment Setup

To illustrate how RBP is used to predict player outcomes, we apply it to predict passing yards per game for NFL quarterbacks with at least 100 pass attempts for the 2021 through 2025 regular seasons.

Our training sample comprises all individual quarterback seasons with at least 100 pass attempts since the 1999 season. We use each quarterback's passing yards per game in the prior season as the current value for predicting his passing yards per game for the upcoming season.

We form our predictions based on the following set of predictive variables.

Exhibit 2: Predictive Variables

Contextual Factors	Prior Season Passing	Prior Season EPA	Prior Season Rushing
Years Played	TD-INT%	EPA/Dropback	Rushing Attempts
Pass Attempts	ANY/A	Interception EPA	Yards/Carry
Height	Success Rate	Air Yards EPA	
Weight	Passer Rating	YAC EPA	
	Yards/Attempt	Sack EPA	
	Yards/Game	Rushing EPA	
	aDOT	Total EPA	
	CPOE		

Results

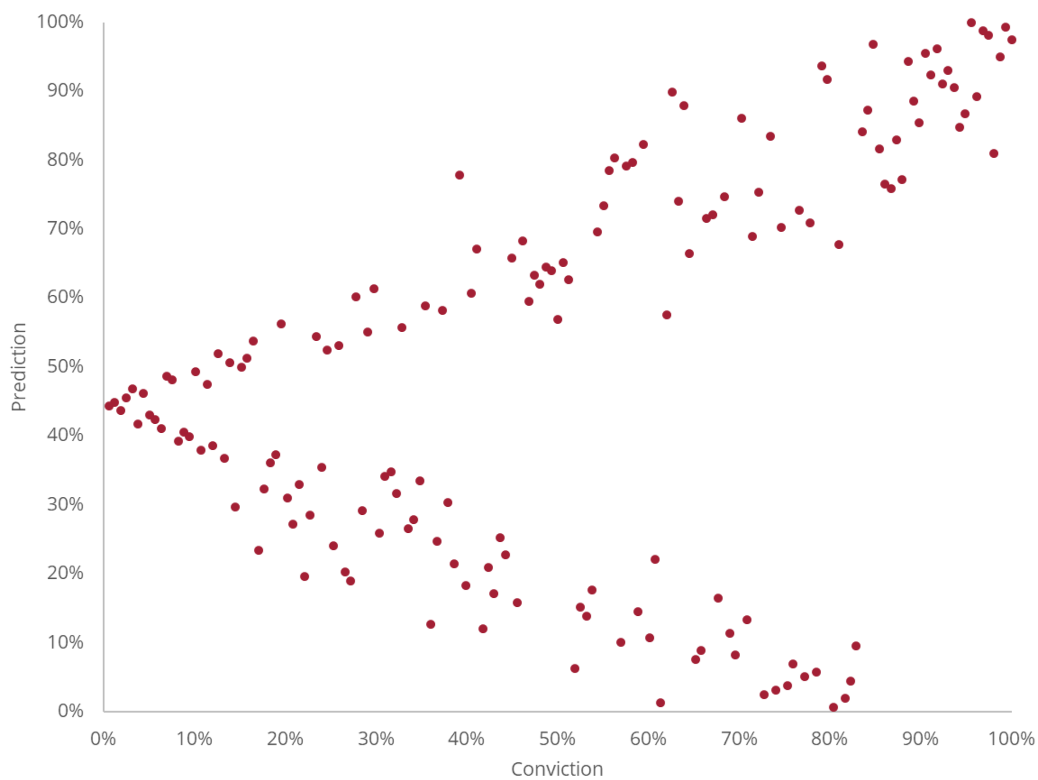
<u>Transparency</u>

Exhibit 3 shows a scatter plot of predicted passing yards per game for quarterbacks for the 2021 through 2025 regular seasons, reported as cross-sectional percentile ranks on the vertical axis, and their corresponding conviction levels based on fit, also reported as cross-sectional percentile ranks, on the horizontal axis. These results yield two key insights. First, conviction varies dramatically from one prediction to the next, even for predictions at similar levels. This underscores the value of fit, which reveals the reliability of each prediction before it is made, thereby enabling teams to view more cautiously predictions that are likely to be less trustworthy. Second, larger magnitude predictions tend be based on stronger patterns leading to higher conviction, while lower conviction predictions based on weaker patterns tend to revert to the mean.

It is important to contrast this prediction-specific detail with the assessment of conviction given by a model. A linear regression model, for example, assigns the same R-squared to all predictions, ignoring the vast differences in reliability across predictions as shown by exhibit 3.

Exhibit 3: Passing Yards per Game Predictions and Convictions

2021 through 2025 Season



In addition to customizing each prediction task to account for a quarterback's specific circumstances, RBP reveals precisely how each previous quarterback informs the prediction. For example, exhibit 4 shows the 10 most important quarterbacks for forming the prediction for Sam Darnold. The product of the relevance weights of these quarterbacks, shown in the third column, and the outcomes shown in the right-most column, along with the weights and outcomes of all the other relevant quarterbacks, gives the prediction of passing yards per game for Darnold. Darnold's prediction ranked at the 69.0th percentile among all quarterbacks for the last five seasons. At the 71.5th percentile, it also had above average conviction. Not surprisingly, the players who were most relevant for forming Soto's prediction had similarly strong outcomes for the seasons in which they were most relevant.

Because RBP gives the identity of the quarterbacks who are most relevant to each prediction, it enables us to observe the characteristics of those quarterbacks that explain their relevance. Exhibit 4, for example, shows which variables best explain the relevant quarterback's similarity to Darnold. It is important to keep in mind, though, that a quarterback's similarity also accounts for covariation across the predictive variables which is not shown here. Nonetheless, this basic information about the quarterbacks who are most relevant to the formation of Darnold's prediction, and which is unobtainable from a model, goes a long way in facilitating dialogue between analytics professionals, coaches, and scouts.

Exhibit 4: Most Relevant Players for Prediction of Sam Darnold's Passing Yards per Game

Sam Darnold (2025) Seahawks, Quarterback						
Prediction Percentile: 69.0%			Conviction Percentile: 71.5%			
10 Most Relevant Players	Season	Weight	Most Similar Characteristics			Yards/Game Ouctome
Russell Wilson	2018	1.895%	Years Played	YAC EPA	Air Yards EPA	256.9
Geno Smith	2022	1.683%	Height	Yards/Game	Rushing EPA	241.6
Deshaun Watson	2018	1.459%	ANY/A	Success Rate	Passer Rating	256.8
Russell Wilson	2019	1.426%	Yards/Game	EPA/Dropback	Yards/Attempt	263.3
Russell Wilson	2020	1.403%	Success Rate	Sack EPA	ANY/A	222.4
Ben Roethlisberger	2007	1.347%	YAC EPA	Yards/Attempt	Passer Rating	207.1
Ryan Tannehill	2019	1.246%	Years Played	Yards/Game	Total EPA	238.7
Aaron Rodgers	2011	1.194%	Weight	Yards/Carry	Air Yards EPA	268.4
Tony Romo	2014	1.150%	Air Yards EPA	Yards/Game	Rushing EPA	221.0
Matthew Stafford	2021	1.147%	Height	aDOT	Passer Rating	231.9

Exhibit 5 presents the same information for Patrick Mahomes. RBP predicted that Mahomes' passing yards per game would rank at the 73.4th percentile with conviction at the 55.1st percentile. Together, exhibits 4 and 5 show how RBP can give similar predictions for two quarterbacks but reveal that one is much more reliable than the other. They also highlight how RBP conditions each prediction on the specific attributes of the quarterback for whom the prediction is made, which explains why each prediction is informed by different quarterbacks. To emphasize this point, it is worth noting that the weights of the most relevant quarterbacks for Darnold's prediction are much smaller than the 10 most relevant players for Mahome's prediction. This difference occurs because there are fewer quarterbacks who are like Mahomes than Darnold; hence the relatively few relevant quarterbacks for Mahomes are weighted more heavily on average than the greater number of relevant quarterbacks for Darnold. This difference also explains why the prediction for Mahomes is less reliable.

Exhibit 5: Most Relevant Players for Prediction of Patrick Mahomes Passing Yards per Game

Patrick Mahomes (2025) Chiefs, Quarterback							
Prediction Percentile: 73.4%			Conviction Percentile: 55.1%				
10 Most Relevant Players	Season	Weight	Most Similar Characteristics			Yards/Game Ouctome
Patrick Mahomes	2021	2.586%	Weight	CPOE	Rushing EPA	308.8
Patrick Mahomes	2022	2.457%	Height	Weight	Rushing EPA	261.4
Patrick Mahomes	2023	2.200%	Height	Weight	CPOE	245.5
Dak Prescott	2023	1.856%	Years Played	Height	Rushing EPA	247.3
Matthew Stafford	2016	1.707%	Years Played	Passer Rating	Success Rate	277.9
Ben Roethlisberger	2020	1.706%	Passer Rating	YAC EPA	Yards/Game	233.8
Christian Ponder	2012	1.683%	Height	Weight	Rushing Attempts	183.1
Kyler Murray	2022	1.584%	Rushing EPA	Air Yards EPA	Sack EPA	224.9
Aaron Rodgers	2020	1.580%	Height	Rushing EPA	Weight	257.2
Drew Brees	2011	1.503%	YAC EPA	Interception EPA	Years Played	323.6

Thus far, we have shown that RBP reveals the specific reliability of each prediction, unlike R-squared which treats the reliability of all predictions the same. And we have shown that RBP precisely quantifies how each previous quarterback informs a current quarterback's prediction and why these previous quarterbacks are informative. This information is unknowable for predictions that come from models. We now turn to RBP's evaluation of predictive variables.

Exhibit 6 shows how each predictive variable contributed to the reliability of each quarterback's prediction of passing yards per game for the 2021 through 2025 seasons, as measured by RBI. The gray bars show the 20th to 80th percentile range of the variables' contributions to reliability across all quarterbacks for whom RBP formed predictions. The lines within the gray bars represent contributions to reliability for the median quarterback. The blue and red diamonds show RBI for Darnold and Mahomes, respectively.

Exhibit 6: Relevance-Based Importance for Passing Yards per Game Predictions for Darnold and Mahomes Relative to All Quarterbacks

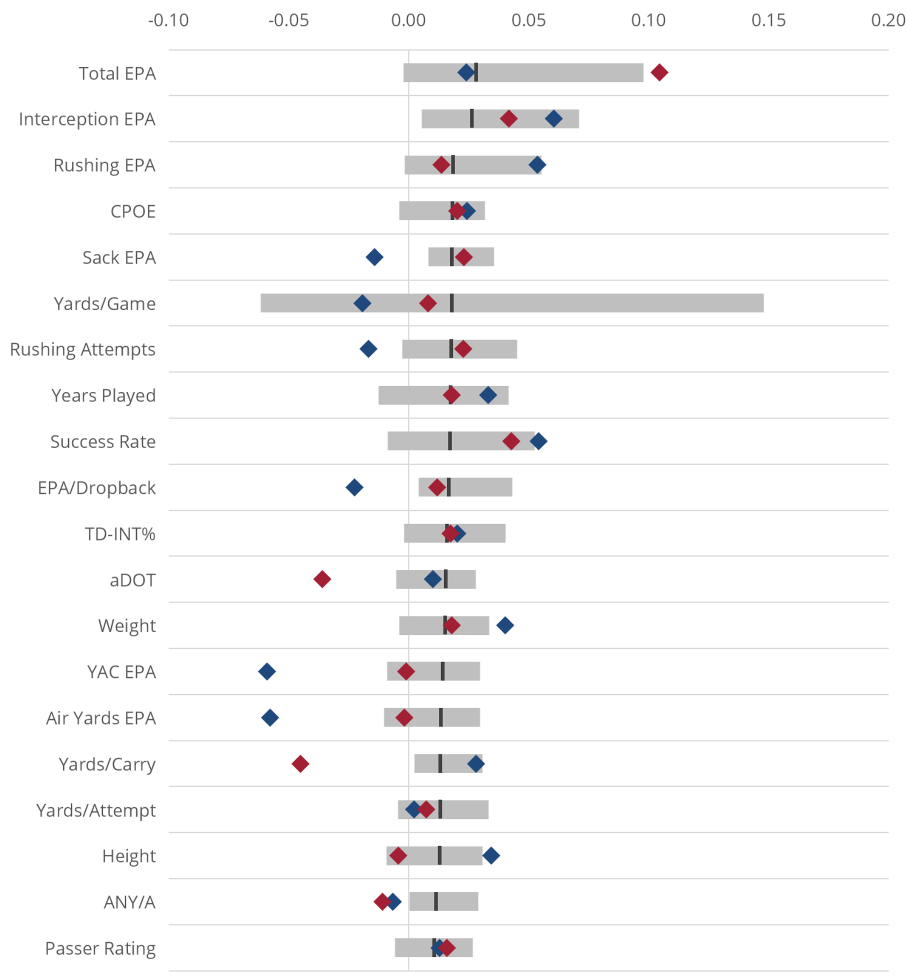
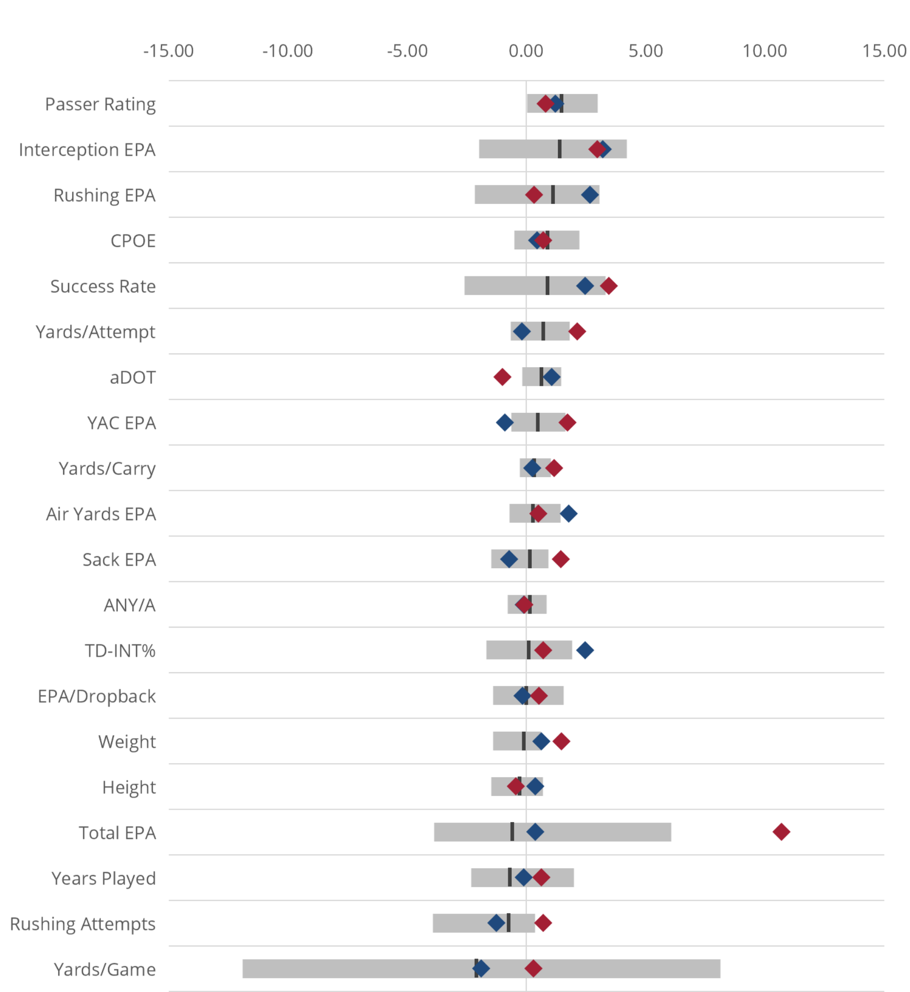


Exhibit 6 reinforces the importance of obtaining prediction-specific information. Whereas a t-statistic would assign the same average importance of a variable to all quarterbacks, exhibit 6 shows that the influence of the predictive variables differs substantially across quarterbacks. For example, Total EPA was extremely important to the reliability of Mahome's prediction but considerably less so for Darnold.

Exhibit 7 shows the contribution of each predictive variable to the passing yards per game (CTP) for the 2021 through 2025 seasons. Again, the gray bars show the 20th to 80th percentile range, the lines within the gray bars represent the median quarterback, the blue diamonds show the contributions of each predictive variable for Sam Darnold's passing yards per game prediction, and the red diamonds show this measure for Patrick Mahomes' prediction.

Exhibit 7: Contribution of Predictive Variables Passing Yards per Game Predictions for Darnold and Mahomes Relative to All Quarterbacks



The key takeaway from exhibit 7 is the huge variation in the contribution of Yards/Game across all the quarterbacks, which was also true for its contribution to prediction reliability. A linear regression model's beta, by contrast, would judge the contribution of each predictive variable to be proportional to the same beta coefficient for all the quarterbacks.

Exhibit 8 compares the quarterback-specific information given by RBP with analogous information that would come from a linear regression model. We should note here that machine learning models have even less transparency than a linear regression model.

Exhibit 8: Transparency of RBP versus Linear Regression Analysis

	RBP	Linear Regression Analysis
Observations	RBP precisely quantifies how each quarterback informs the prediction.	Linear regression analysis offes no visibiltiy into the influence of individual quarterbacks on the formation of the prediction.
Conviction	Fit measures the unique reliability of each prediction, which across all predictions, aggregates to R-squared.	R-squared only measures a model's reliability. It assigns the same level of conviction to all predictions.
Contribution to prediction reliability	RBI measures how each preditive variable uniquely contributes to the reliability of each prediction.	A t-statistic assumes that each variable is equally important to all predictions.
Contribution to prediction value	CTP measures the contribution of each predictive variable to the value of each prediction.	Linear regression analysis judges the contribution of each predictive variable to be proportional to the same beta coefficient for all quarterbacks.

Efficacy

We have so far given compelling evidence that RBP offers remarkable visibility into the formation of individual predictions. Additionally, RBP measures the unique reliability of each prediction, the contribution of each predictive variable to the reliability of each prediction, and the contribution of each predictive variable to the magnitude of each prediction. RBP's ability to yield this nuanced information extends far beyond the capabilities of linear regression analysis and other model-based approaches to prediction. We now turn our attention to RBP's predictive efficacy.

Exhibit 9 shows the realized passing yards per game for quarterbacks with 100 or more passing attempts for the 2021 through 2025 seasons. The first row shows the outcomes for quarterbacks who were predicted to be in the top half of all quarterbacks, while the middle row shows the outcomes for quarterbacks predicted to be in the bottom half of all quarterbacks. The third row gives the spread between the outcomes of quarterbacks who were predicted to be in the top half set against those predicted to be in the bottom half. It, therefore, serves as a measure of prediction efficacy.

The first column reports the results given by RBP across all predictions, including those predictions known in advance to be less reliable. The second column shows the results given by RBP for those predictions judged in advance to be the 50% most reliable predictions based on fit. And the final column shows the results for those predictions anticipated by fit to be the 50% least reliable. These two columns demonstrate that RBP can effectively separate trustworthy predictions from those that are less reliable. By contrast, model-based approaches to prediction give no visibility into which predictions to trust and which to view with skepticism. Given a spread of 70 for the trustworthy predictions versus only 14 for the doubtful predictions, RBP's ability to anticipate each prediction's specific reliability constitutes a huge advantage over model-based approaches to prediction.

Exhibit 9: Realized Passing Yards per Game for Players Predicted to Outperform Versus Players Predicted to Underperform

	RBP Predictions (All)	RBP Predictions (High Conviction)	RBP Predictions (Low Conviction)
Full sample average	223		
High prediction	244	256	231
Low prediction	201	186	217
Spread	42	70	14

Summary

We described a new approach for predicting performance outcomes for NFL quarterbacks called relevance-based prediction. We illustrated RBP by predicting passing yards per game for quarterbacks with 100 or more passing attempts for the 2021 through 2025 seasons.

Our analysis highlighted the extraordinary transparency of RBP. We reported how specific quarterbacks contribute to the formation of individual predictions, which is almost always unobservable for predictions generated by models. We reported the specific reliability of individual predictions in contrast to R-squared which only gives a model’s average reliability. We showed the contribution of each predictive variable to the reliability of individual predictions in contrast to a t-statistic which only measures a variable’s average importance. And we reported the unique contribution of each predictive variable to the value of each prediction in contrast to a linear regression equation’s beta which assumes a predictive variable’s contribution to the value of a prediction is proportional to the same parameter value across all predictions.

Finally, we showed that RBP successfully distinguished in advance quarterbacks who produced more favorable outcomes from those who produced less favorable outcomes, and we demonstrated that, unlike model-based approaches to prediction, RBP could distinguish in advance which predictions to trust and which to discard or treat with caution.

References

[1] Czasonis, Megan, Mark Kritzman, and David Turkington. 2022a. "Relevance." *The Journal of Investment Management*, 20 (1).

[2] Czasonis, Megan, Mark Kritzman, and David Turkington. 2022b. *Prediction Revisited: The Importance of Observation*. New Jersey: John S. Wiley & Sons.

[3] Czasonis, Megan, Mark Kritzman, and David Turkington. 2023. "Relevance-Based Prediction: A Transparent and Adaptive Alternative to Machine Learning." *The Journal of Financial Data Analysis,* 5, (1).

[4] Czasonis, Megan, Mark Kritzman, and David Turkington. 2024. "The Virtue of Transparency: How to Maximize the Utility of Data Without Overfitting." *The Journal of Financial Data Science*, 7 (2).

[5] Czasonis, Megan, Mark Kritzman, and David Turkington. 2025a. "Prediction with Incomplete Information." *The Journal of Financial Data Science*, 7 (4).

[6] Czasonis, Megan, Mark Kritzman, and David Turkington. 2025b. "Relevance-Based Importance: A Comprehensive Measure of Variable Importance in Prediction." *The Journal of Portfolio Management*, 51 (9).

[7] Czasonis, Megan, Mark Kritzman, and David Turkington. 2025c. "A Transparent Alternative to Neural Networks with an Application to Predicting Volatility." *The Journal of Investment Management*, 23 (3).

[8] Czasonis, Megan, Miles Kee, Mark Kritzman, and David Turkington. 2025d. "Prediction with Transparency: Offensive Value in Baseball." *MIT Sloan Working Paper*, No. 7338-25

[9] Mahalanobis, Prasanta Chandra. 1936. "On the Generalised Distance in Statistics." *Proceedings of the National Institute of Sciences of India* 2, no. 1: 49–55.

[10] Shannon, Claude. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical Journal*, 27 (July, October): 379–423, 623–656.

[11] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. “Attention is All you Need.” In *Advances in Neural Information Processing Systems* (vol. 30).

---

[1] The descriptions of the features of RBP and variable importance follow closely language used from Czasonis, Kritzman, and Turkington (2022a, 2022b, 2023, 2024, 2025a, 2025b, 2025c, and 2025d), but they are modified to fit the context of the current discussion.

[2] We multiply by one half because the average squared distances between pairs of players is twice as large as the players' average squared differences from the average of all players.

[3] See Czasonis, Kritzman, and Turkington (2022b) for proof of this result.

[4] See Czasonis, Kritzman, and Turkington (2023) for proof of this result.

[5] See Czasonis, Kritzman, and Turkington (2022b) for proof of this result.

[6] The formula we present gives a useful indication of the contribution of each variable, but it does not constitute an exact decomposition of the grid prediction value. To calculate an exact decomposition would require forming new composite grid predictions for every subset of variables.