# PREDICTION WITH CONVICTION:
# AN APPLICATION OF RELEVANCE-BASED PREDICTION TO THE NBA

THIS VERSION SEPTEMBER 17, 2025

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

**Cel Kulasekaran** is a Founding Partner of Cambridge Sports Analytics in Cambridge, MA.
100 Main Street, Cambridge MA, 02142
ckulasekaran@csanalytics.io

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

Abstract

The authors describe a new prediction system, called relevance-based prediction (RBP), for predicting player performance for NBA draft prospects based on the outcomes of previous NBA players. This approach rests on a statistical concept called relevance, which gives a mathematically precise and theoretically justified measure of the importance of a previous player to a prediction. The authors also describe fit, which gives advance guidance about the reliability of a specific prediction. And they show how fit, together with asymmetry, focuses each prediction on the combinations of predictive variables and previous players that are most effective for that prediction task. The authors argue that RBP addresses complexities that are beyond the reach of conventional prediction models, but in a way that is more transparent, more flexible, and more theoretically justified than widely used machine learning algorithms.

# PREDICTION WITH CONVICTION:
# AN APPLICATION OF RELEVANCE-BASED PREDICTION TO THE NBA

We propose a new prediction system, called relevance-based prediction (RBP), for predicting player outcomes. RBP forms a prediction as a weighted average of observed outcomes in which the weights are based on a rigorously defined and theoretically justified statistic called relevance. Unlike predictive models such as linear regression analysis or machine learning, which work by estimating model parameters and then applying those parameters to new tasks, RBP is fundamentally model-free. It works by evaluating patterns in the relationship between outcomes and predictive variables given the specific circumstances of each prediction task.

RBP has several significant advantages compared to model-based prediction.

- RBP is theoretically justified by information theory, the Central Limit Theorem, the Mahalanobis distance, and surprising mathematical convergences.
- RBP is prediction specific. It tailors the choice of players and predictive variables to each individual prediction task.
- RBP is fully transparent. It reveals precisely how each player informs an individual prediction and how each predictive variable contributes to the reliability of the prediction.
- RBP reveals the reliability of each prediction before it is made, thereby enabling analysts to view more cautiously predictions that are likely to be less trustworthy.
- RBP is less vulnerable to overfitting small samples because it proceeds task by task, its transparency reveals the unique circumstances of each player who informs the prediction, and it optimally balances the stronger patterns that exist within a more focused sample with the increase in noise that comes with intentionally shrinking the sample.

- RBP is resilient to missing information.[1] It explicitly accounts for the relative importance of missing information in forming a prediction and evaluating its reliability, and it retains information that model-based prediction would exclude.

As has been well documented in the finance and data science literature, RBP extracts as much information from complex datasets as machine learning models but more efficiently and with full transparency.[2] Our purpose in this paper, however, is to highlight its application to predictive sports analytics. We proceed by first describing the key features of RBP conceptually and mathematically. We then illustrate RBP by showing how it would have predicted VORP (value over replacement player)[3] of NBA players during their rookie seasons, based on certain attributes of these players and their pre-NBA basketball performance, as well as attributes and performance of NBA players who came before them, and we compare our predictions to the draft prospects' actual outcomes in the NBA. 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 technique that forms a prediction as a relevance-weighted average of observed outcomes in which relevance has a precise statistical meaning. Although RBP gives the same prediction as linear regression analysis if it is applied across all players, it usually gives a more reliable prediction if it is applied to a subset of relevant players. When RBP is applied to a subset of relevant players, it is called partial sample regression. RBP

also depends crucially on fit, which measures the average alignment of relevance and outcomes across all pairs of players that go into a prediction task. Fit assesses the expected reliability of individual predictions before they are rendered. The final feature of RBP is grid prediction, which forms a composite prediction as a reliability-weighted average of many predictions given by different combinations of players and predictive variables.

Relevance

Relevance is a statistical measure of the importance of previous NBA players to forming a prediction for an NBA draft prospect given a chosen set of predictive variables. It is composed of two components, similarity and informativeness, as shown in Equation 1.

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

In Equation 1, similarity and informativeness are computed as Mahalanobis distances (Mahalanobis 1936) rather than absolute distances or Euclidean distances.

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

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

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

In Equations 1 through 4, $x_i$ is a row vector of the values of the predictive variables for a past player, $x_t$ is a row vector of the values of the predictive variables for the prospective player, $\bar{x}$ is a vector of the average values of the predictive variables for all previous players in the sample, $\Omega^{-1}$ is the inverse covariance matrix of the values of the predictive variables for all previous players, and $'$ denotes matrix transpose.

The vector $(x_i - x_t)$ measures how different a previous player is from the prospect, whereas the vector $(x_i - \bar{x})$ measures how different he is from average, and $(x_t - \bar{x})$ measures how different the prospect is from average. By multiplying these vectors by the inverse of the covariance matrix, we capture the correlation of the attributes of the previous players. Also, this calculation implicitly standardizes the differences by dividing them by variance. By multiplying the product by the transpose of the vector difference we consolidate the outcome into a single number, which represents the covariance-adjusted distance between the two vectors.

Notice that in the formula for similarity we multiply the Mahalanobis distance of a previous player from the prospect by negative one half. The negative sign converts a measure of difference into a measure of similarity. 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. When we measure informativeness, we retain its positive sign, and we need not multiply by one half. By measuring informativeness as a difference from average, we are recognizing that unusual players contain more information than typical players. Intuitively, this occurs because the outcomes for an unusual player are likely to reveal underlying relationships to his personal attributes and circumstances, whereas outcomes for highly typical players are likely to contain more noise and less information. Finally, note that we measure the unusualness of the prospect. We do so to center our measure of relevance on zero. All else being equal, previous players who are like the prospect but different from the average of all previous players are more relevant to a prediction than those who are not.

This definition of relevance is not arbitrary. We know from information theory that the information contained in an observation is the negative logarithm of its likelihood (Shannon 1948). We also know from the Central Limit Theorem that the relative likelihood of an observation from a multivariate normal distribution is proportional to the exponential of a negative Mahalanobis distance. Therefore, the information contained in a point on a multivariate normal distribution is proportional to a Mahalanobis distance.

We can also justify the nonarbitrariness of relevance by considering a limiting case of the predictions it yields. RBP forms a prediction as a weighted average of prior player outcomes for $Y$.

$$\hat{y}_t = \sum_{i=1}^{N} w_{it} y_i \tag{5}$$

If we define weights in terms of relevance as follows, which admits the relevance-weighted average of every prior player outcome in the observed data sample, the result is precisely equivalent to the prediction that results from linear regression analysis. [4]

$$w_{it,linear} = \frac{1}{N} + \frac{1}{N-1} r_{it} \tag{6}$$

Owing to this equivalence, the theoretical justification given by Gauss for linear regression analysis applies as well to RBP.[5] In most cases, however, we can produce a more reliable prediction by taking a relevance-weighted average of a subset of relevant players, which is called partial sample regression. Partial sample regression censors the influence of past players who are less relevant than a chosen threshold, which leads to the following definition of prediction weights.

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

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

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

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. It is important to note that $w_{it,retained}$ depends crucially on the prediction circumstances $x_t$. Relevance is reassessed for each prediction circumstance which further affects the identification of the retained subsample and introduces nonlinear conditional dependence of the prediction $\hat{y}_t$ on the prediction circumstances $x_t$ . The scaling factor $\lambda^2$ compensates for a bias that would otherwise result from relying on a small subsample of highly relevant players. In the case of linear regression analysis $n = N$ and $\lambda^2 = 1$. Lastly, note that the regression weights always sum to 1.[6]

Fit

Fit is a critical component of RBP. It reveals how much confidence we should have in a specific prediction task, separately from the confidence we have in the overall prediction system. In addition, fit provides a principled way to evaluate the relative merits of alternative calibrations for each prediction task.

Consider, for example, a pair of previous players who are used, in part, to form the prediction of an outcome for a prospect. Each previous player has a relevance weight and an

outcome. We are interested in the alignment of the relevance weights of the two previous players with their outcomes. But we must standardize them by subtracting the average value and dividing by standard deviation – in essence, converting them to z-scores. We then measure their alignment by taking the product of the standardized values. If this product is positive, their relevance is aligned with their outcomes, and the larger the product, the stronger the alignment. We perform this calculation for every pair of previous players in our sample. We should also note that all the formulas we have thus far considered for the relevance weights rely only on the $x_i s$, the $x_t s$, and the $\bar{x}s$. They do not make use of any of the information from previous player outcomes. To determine fit, however, we must consider outcomes (the $y_i$s).

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

Equation 11 intuitively describes fit as the squared correlation of relevance weights and outcomes, which conceptually matches the notion of the conventional R-squared statistic. As we soon show, this connection of fit to R-squared is critically important.

$$fit_t = \rho(w_t, y)^2 \tag{11}$$

Although we compute fit from the full sample of players, the weights that determine fit vary with the threshold we choose to define the relevant subsample. As we focus the subsample on players who are more relevant, we should expect the fit of the subsample to increase, but we should also expect more noise as we shrink the number of players. The fit across pairs of all players in the full sample implicitly captures this tradeoff between subsample fit and noise by overweighting players who are more relevant and underweighting players who are less relevant accordingly.

Like relevance, fit is not arbitrary. In the case of linear regression analysis with $n = N$, the informativeness-weighted average fit across all prediction tasks in the observed sample equals R-squared.[7]

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

This convergence of fit to R-squared reveals an intriguing insight. R-squared is the result of some good predictions, some average predictions, and some bad predictions; that is, some predictions with high fit, some with average fit, and some with low fit. R-squared reveals the average reliability of a prediction model. It reveals much less about the reliability of specific prediction tasks, which can vary substantially. Fit is much more nuanced. It gauges the reliability of a specific prediction task in a non-arbitrary way, as demonstrated by its convergence to R-squared. Fit is the fundamental building block of R-squared. To compute fit, we must know the weight of each observation 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 observations 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 analysts to discard or view with greater caution predictions that are foreseen to be unreliable.

Grid Prediction

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 left unanswered the question of how to determine the threshold

for the subsample of relevant players. We have only noted that a partial sample regression prediction depends on the choice of a parameter, $r^*$, which is the censoring threshold for relevance. In addition, we have implicitly assumed up to this point that the full menu of predictive variables is used to measure relevance and form a partial sample prediction. However, it is possible that a subset of the predictive variables will render a better prediction for a specific prediction task. The efficacy of previous players for a given prediction task depends on the predictive variables, and the efficacy of the predictive variables depends on the players. These choices are codependent. We, therefore, turn to the last key feature of RBP, which is grid prediction. But before we show how to form predictions that consider a range of alternative calibrations, we must first describe an enhanced version of fit called adjusted fit.

Partial sample prediction is more effective to the extent there is strong alignment between relevance and outcomes, as measured by fit. It is also more effective to the extent there is asymmetry between the fit of the retained subsample of previous players and the fit of the censored players. In the presence of asymmetry, we trust the more relevant sample on principle. In the absence of asymmetry, the full sample relationship is linear, and linear regression analysis will suffice. Therefore, to compare properly the efficacy of two predictions formed from different values of $r^*$, we need a way to measure not only fit but asymmetry.

We measure asymmetry between the fit of the retained and censored subsamples as shown by Equation 13.

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

The $(+)$ superscript designates weights formed from the retained subsample of players while the $(-)$ superscript designates weights formed from the complementary sample of censored players. Asymmetry recognizes the benefit of censoring non-relevant players that contradict the predictive relationships that exist among the relevant observations. This assessment also inherently considers the relative sample sizes of the complementary groups

To calculate adjusted fit, we add asymmetry to fit and multiply this sum by $K$, the number of predictive variables, as shown by Equation 14.

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

Multiplication by the number of predictive variables allows us to compare predictions based on different numbers of predictive variables. It corrects a bias that would otherwise occur, whereby adding a pure noise variable decreases fit in proportion to the increase in the number of variables, even if the predictions themselves do not change (consider, for example, the case of a full sample linear regression analysis with a large sample of players). Another way to view the intuition for $K$ is that we are more likely to observe a spurious relationship from weights based on any one variable in isolation than we are based on a collection of many variables.

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 could consider every possible calibration that combines a choice of $r^*$ with a choice of a subset of variables and select the prediction with the greatest reliability as measured by adjusted fit. It is critical to remember that the assessment of reliability using adjusted fit is

made before the prediction is rendered and the subsequent outcome is known. And it is also critical to remember that the assessment of reliability is specific to the prediction task.

Instead of selecting one optimal calibration for a given prediction task, it may be more prudent to compute a composite prediction as a reliability-weighted average of the predictions from all possible calibrations. Equation 15 defines reliability weights, $\psi_\theta$, as the adjusted fit for a parameter calibration, $\theta$, divided by the sum of all adjusted fits across all parameter calibrations.

$$\psi_\theta = \frac{adjusted\ fit_\theta}{\sum_{\tilde{\theta}} adjusted\ fit_{\tilde{\theta}}} \tag{15}$$

Equation 16 describes the composite prediction.

$$\hat{y}_{t,grid} = \sum_\theta \psi_\theta \hat{y}_{t,\theta} \tag{16}$$

Figure 1 gives a visual representation of grid prediction. 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 $\theta$; 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 Figure 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 $\psi_\theta$ we apply to the calibration-specific predictions to form the composite prediction. The values in the grid are specific to each 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%).

Figure 1: Grid Prediction – Illustrative Example

Variable Combinations

| Observation 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

Note that each cell's prediction is a linear function of player observations, and the grid prediction is a linear function of each cell's prediction. Therefore, we can express the grid prediction in terms of composite weights applied to each player, as shown in Equation 17. Composite weights are important because they preserve the transparency of how each previous player contributes to the current prediction task, and they allow us to calculate fit from composite weights as a final gauge of the grid prediction's reliability.

$$w_{it,grid} = \sum_{\theta} \psi_{\theta} w_{it,\theta} \tag{17}$$

The prediction grid also yields a comprehensive measure of how important each variable is to the reliability of the current prediction. This measure is called relevance-based importance (RBI).[8] As shown by Equation 18, $RBI_{tk}$ for prediction $t$ and variable $k$ is computed as the weighted average adjusted fit for grid cells that contain $k$ (for which the variable censoring indicator $\Delta_k(\theta) = 1$) minus the weighted average adjusted fit for cells that do not contain $k$ (for which $\Delta_k(\theta) = 0$). We express RBI as a sum over all grid cells $\theta$.

$$RBI_{tk} = \sum_{\theta} \alpha_{\theta} \frac{\Delta_k(\theta)(adjusted\ fit_{t\theta}) - (1-\Delta_k(\theta))(adjusted\ fit_{t\theta})}{\sum_{\tilde{\theta}} \Delta_k(\tilde{\theta})} \quad (18)$$

The term $\sum_{\tilde{\theta}} \Delta_k(\tilde{\theta})$ counts the number of cells that include variable $k$. For a grid that includes every variable combination, this number is nearly equal to the number of cells that do not include variable $k$, but the counts are not identical unless we include a column in the grid for predictions that do not use any of the $X$ variables (for which adjusted fit is always zero). Thus, we divide by the number of cells that include variable $k$ regardless of whether a given cell contains $k$ or not.

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.

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 these player 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.

**Application of Relevance-Based Prediction for NBA Outcomes**

To illustrate how RBP is used to predict player outcomes, we apply it to predict VORP (value over replacement player) during their NBA rookie years for players drafted in 2022, 2023, and 2024. We chose to predict VORP because it reflects a variety of ways in which a player affects scoring and therefore has the potential to incorporate hidden complexities. Having said that, we wish to emphasize that RBP can be applied to predict any outcome for any player.

Our full data sample comprises 468 players who were drafted from 2011 through 2024, excluding 2019 and 2020, from Division I U.S. colleges with at least one season in the NBA. We excluded players from 2019 and 2020 to avoid distortions that might have occurred from the effect of COVID on both the collegiate and NBA player statistics. For each player in the 2022, 2023, and 2024 drafts and for each previously drafted player, we collect data in four categories: physical attributes, individual college performance, team performance in college, and team performance in the NBA. We chose these predictive variables merely to illustrate RBP. We do not have expertise in determining the most effective predictive variables. For each prediction, we use training data from previously drafted players that would have been available at the time of that year's draft.

Prediction task:

- Rookie year VORP for 2022, 2023, and 2024 draft cohorts

Training sample:

- Players drafted from 2011 through 2023 who were drafted prior to the draft class that is currently being predicted, excluding 2019 and 2020, from Division I colleges with at least one NBA season

Predictive variables:

- Physical attributes
  - Height
  - Weight
- College performance (final college season)
  - Net rating
  - True shooting percentage
  - 3-point percentage
  - Free throw rate
  - Assists percentage
  - Offensive box plus minus (BPM)
  - Defensive box plus minus (BPM)
  - Minutes per game
- Non-player factors – College
  - School's conference winning percentage (final college season)
  - Number of players from school drafted into the NBA (prior 10 years)
- Non-player factors – NBA team (prior season)
  - Win percentage
  - Average point spread

As we discussed previously, grid prediction considers many subsamples of previously drafted players and subsets of predictive variables for each individual prediction task. The information from every cell in the grid is aggregated to form one composite vector of weights across all previously drafted players. These weights directly determine the prediction: it equals the weighted average of player outcomes. The weights also contain other important information. For illustrative purposes, let us consider the VORP predictions for the three draft cohorts. For the players we predicted, Figures 2 through 4 show scatter plots of predicted rookie year VORPs, 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 reveal 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 analysts 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. This underscores RBP's model-free approach to prediction. It works by assessing patterns in past players specific to the circumstances of each prediction task, thereby calibrating each individual prediction one task at a time using a rigorous evaluation routine.

Figure 2: VORP Predictions and Convictions for 2022 Draft Cohort

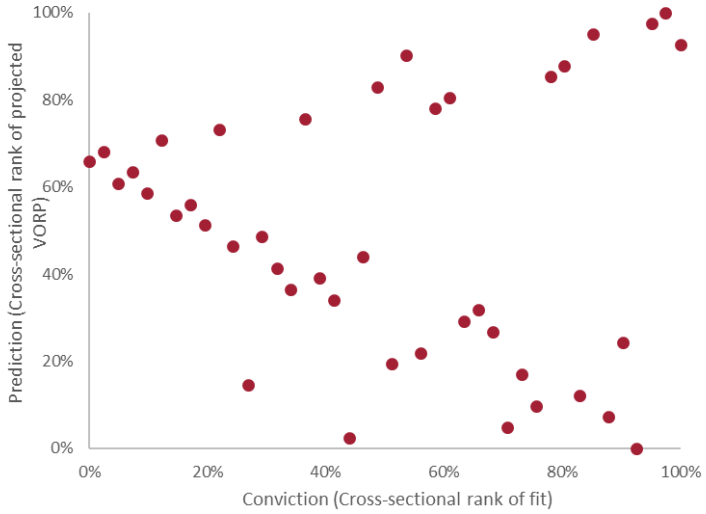


Figure 3: VORP Predictions and Convictions for 2023 Draft Cohort

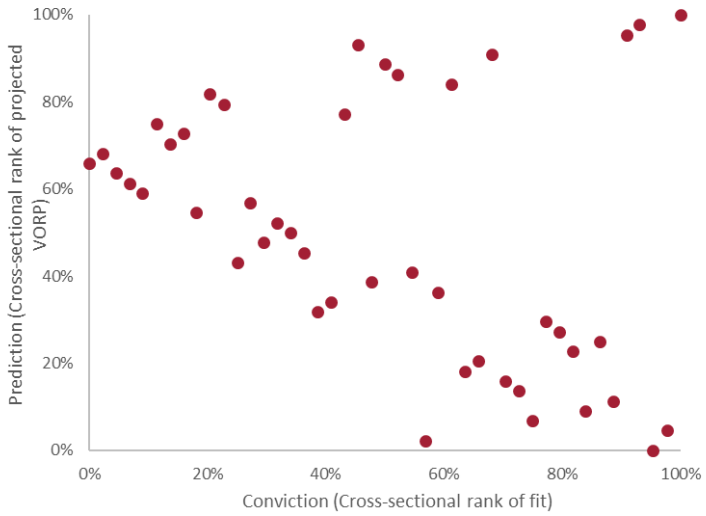
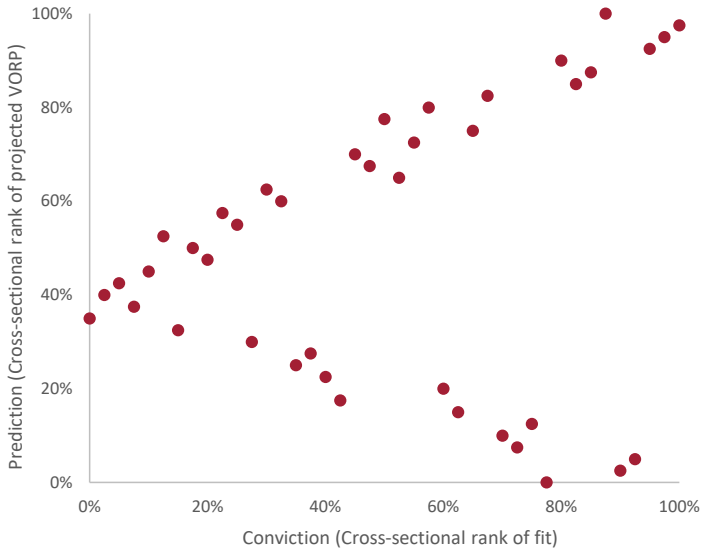



Figure 4: VORP Predictions and Convictions for 2024 Draft Cohort




In addition to customizing each prediction task to account for a draft prospect's unique circumstances, RBP reveals precisely how each previously drafted player informs the prediction. For example, Figure 5 shows the ten most important players for forming the VORP predictions for Zach Edey and Reed Sheppard. Edey's predicted VORP, 0.68, was the highest (100$^{th}$

percentile) among all players in his draft cohort. At the 88$^{th}$ percentile, it also had high conviction. It is affirming to note that RBP successfully identified highly relevant players with similarly strong rookie season performance as what subsequently occurred for Edey. In comparison, Reed Sheppard's predicted VORP, 0.26, was more moderate (80$^{th}$ percentile) along with its conviction (58$^{th}$ percentile). Intuitively, we see that the most relevant players for Sheppard's prediction generally had weaker rookie year performance than those for Edey's prediction. Interestingly, RBP identified three fellow University of Kentucky alums—Isaiah Jackson, Cason Wallace, and TyTy Washington—among the most relevant players for Sheppard. The main takeaway from Figure 5, though, is the extraordinary level of transparency that RBP affords, which is critical for facilitating dialogue between analytics professionals, coaches, and scouts.

Figure 5: Most Important Players for Zach Edey and Reed Sheppard

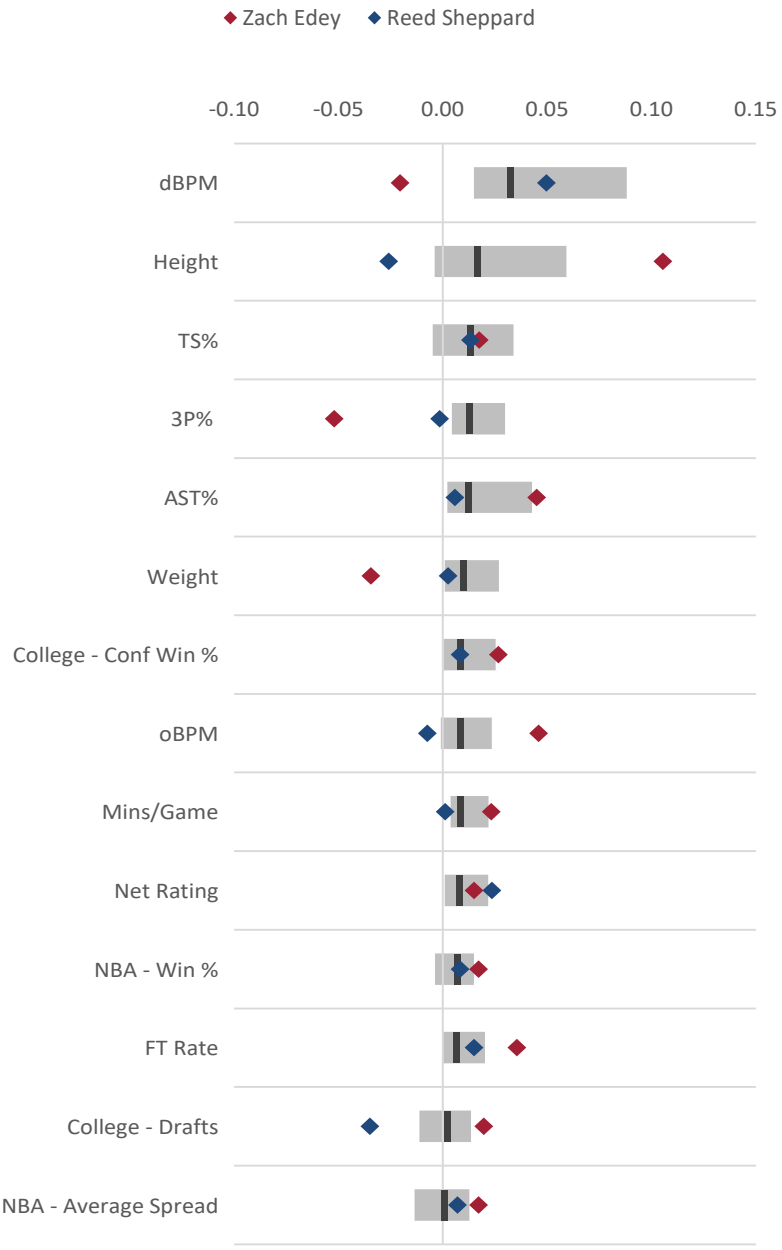
| **Zach Edey** | | | | | |
|---|---|---|---|---|---|
| Purdue | | | | | |
| Prediction (VORP): | 0.68 | | | | |
| Prediction Percentile: | 100% | | | | |
| Conviction Percentile: | 88% | | | | |
| Top 10 Most Important Players (Highest Prediction Weights) | | | | | |
| Prior Player | Weight | Most Similar Variables | | | VORP |
| Luka Garza | 4.6% | Mins/Game | AST% | NBA Win% | -0.1 |
| Frank Kaminsky | 4.4% | College Conf Win% | College Drafts | Net Rating | 0.5 |
| Ben Simmons | 4.1% | dBPM | FT Rate | College Drafts | 4.5 |
| Deandre Ayton | 4.1% | TS% | College Conf Win% | Mins/Game | 1.2 |
| Kelly Olynyk | 3.8% | College Drafts | dBPM | AST% | 0.3 |
| Joel Embiid | 3.6% | TS% | FT Rate | AST% | 1.3 |
| Marvin Bagley | 3.5% | NBA Avg Spread | NBA Win% | Net Rating | 0.4 |
| Cody Zeller | 3.4% | dBPM | Net Rating | College Drafts | 0.0 |
| Cameron Bairstow | 3.2% | College Drafts | AST% | Mins/Game | -0.1 |
| Jared Sullinger | 3.2% | Net Rating | dBPM | Mins/Game | 0.0 |

| **Reed Sheppard** | | | | | |
|---|---|---|---|---|---|
| University of Kentucky | | | | | |
| Prediction (VORP): | 0.26 | | | | |
| Prediction Percentile: | 80% | | | | |
| Conviction Percentile: | 58% | | | | |
| Top 10 Most Important Players (Highest Prediction Weights) | | | | | |
| Prior Player | Weight | Most Similar Variables | | | VORP |
| Dereck Lively | 3.6% | NBA Avg Spread | Net Rating | dBPM | 0.8 |
| Mark Williams | 3.2% | NBA Avg Spread | NBA Win% | FT Rate | 0.4 |
| Shai Gilgeous-Alexander | 2.9% | oBPM | NBA Avg Spread | dBPM | 0.8 |
| Lonzo Ball | 2.7% | Weight | FT Rate | College Conf Win% | 1.3 |
| Isaiah Jackson | 2.6% | dBPM | NBA Avg Spread | Net Rating | 0.0 |
| Cason Wallace | 2.6% | College Drafts | College Conf Win% | AST% | 0.9 |
| Jalen Johnson | 2.5% | Net Rating | NBA Avg Spread | AST% | -0.1 |
| AJ Griffin | 2.3% | NBA Avg Spread | oBPM | NBA Win% | 0.4 |
| TyTy Washington | 2.3% | Height | College Conf Win% | Mins/Game | -0.5 |
| Chet Holmgren | 2.3% | TS% | Weight | oBPM | 3.3 |

Figure 6 gives further evidence of RBP's transparency and the importance of prediction-specific information. It shows how each predictive variable contributed to the reliability of each VORP prediction for the 2024 draft cohort, based on RBI which we described earlier. The gray

bars show the 20th to 80th percentile range of variable importance across all 42 players. The lines within the gray bars represent the median player. The red diamonds show the importance of each predictive variable for Zach Edey, while the blue diamonds show this measure for Reed Sheppard. Figure 6 shows remarkable differences in the importance of the predictive variables across these two players. For example, height was by far the most important variable for Edey's prediction, while it contributed adversely to the reliability of Sheppard's prediction. This suggests that past players with similar statures to Edey (who, at 7 feet 4 inches, is the tallest player in our sample) had relatively consistent rookie year VORPs. Therefore, including height as a predictive variable contributed positively to the reliability of his prediction. However, for Sheppard, including height harmed the reliability of his prediction, indicating inconsistent performance for past players with similar heights. The opposite is true for defensive BPM. It was the most important variable for Sheppard's prediction, but unimportant to Edey's prediction.

Figure 6: Variable Importance for VORP Predictions for 2024 Draft Cohort




Next, we show how well RBP performed out of sample using the same set of predictive variables. Figure 7 shows the average rookie season VORP for players drafted in 2022, 2023, and 2024. The first set of rows shows the actual outcomes for players who were predicted to be in the top half of all players and those predicted to be in the bottom half of all players, based on all RBP predictions for a given draft cohort. The second and third set of rows show the same

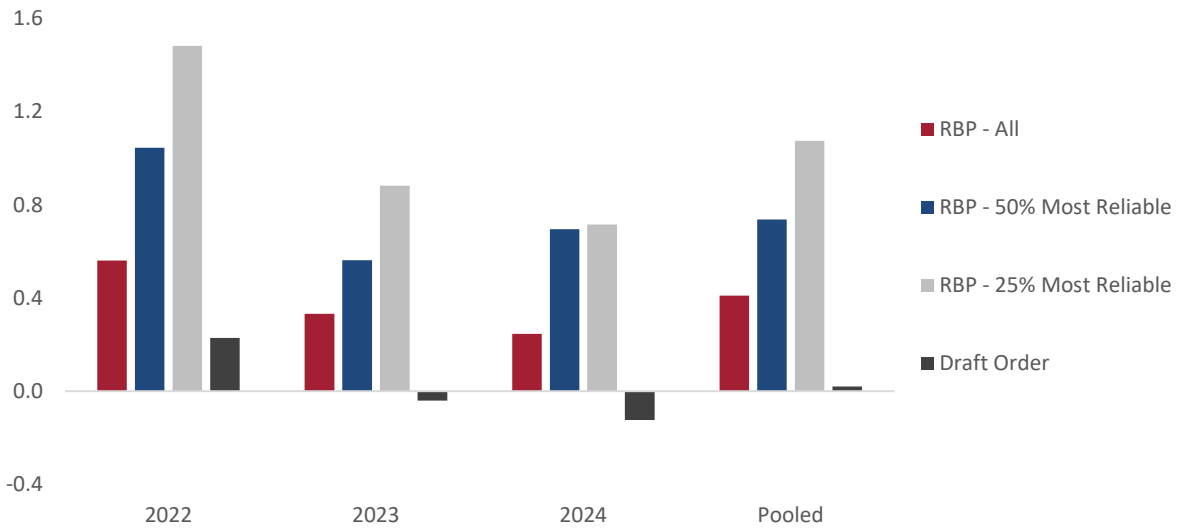
comparisons, but only for those predictions foreseen to be among the 50% and 75% most reliable, respectively, as indicated by fit. The final set of rows compare the VORP outcomes for the players who were among the first 50% to be drafted with those who were among the last 50% to be drafted.

Figure 7: Average Actual Rookie Year VORPs

| | | 2022 | 2023 | 2024 | Pooled |
|---|---|---|---|---|---|
| RBP - All | High prediction | 0.23 | 0.07 | -0.04 | 0.11 |
| | Low prediction | -0.33 | -0.26 | -0.29 | -0.30 |
| RBP - 50% most reliable | High prediction | 0.57 | 0.21 | 0.14 | 0.32 |
| | Low prediction | -0.47 | -0.35 | -0.56 | -0.42 |
| RBP - 25% most reliable | High prediction | 0.97 | 0.38 | 0.22 | 0.58 |
| | Low prediction | -0.52 | -0.50 | -0.50 | -0.50 |
| Draft order | High prediction | 0.07 | -0.11 | -0.22 | -0.09 |
| | Low prediction | -0.16 | -0.07 | -0.10 | -0.11 |

To facilitate comparison, Figure 8 summarizes the spread in realized VORP outcomes for high minus low predictions for the four groups of predictions.

Figure 8: Spread in Average Actual Rookie Year VORPs for High versus Low Predictions



The main takeaway from Figures 7 and 8 is that RBP successively distinguished in advance players with more favorable VORP outcomes from those with less favorable outcomes, especially for predictions foreseen to be more reliable. Moreover, it delineated more favorable outcomes from less favorable outcomes more effectively than the order in which the players were drafted, though it is important to acknowledge that many other factors contribute to the order of the draft.

**Summary**

We described a new approach for predicting player outcomes called relevance-based prediction. RBP forms predictions as weighted averages of past outcomes in which the weights are based on the relevance of previous players, measured in a mathematically precise and theoretically justified way.

Then we described a measure of prediction-specific fit, which indicates the specific reliability of each individual prediction task. R-squared, by comparison, measures only the average reliability of a prediction model. We showed that fit converges to R-squared in the case of linear regression analysis when aggregated properly across all prediction tasks. And of critical importance, we showed that fit enables us to discard, or consider more cautiously, predictions that are foreseen to be less reliable.

Next, we introduced grid prediction, which uses fit to precisely blend the predictions that result from different combinations of players and predictive variables. Crucially, the blend

places greater emphasis on players and variables that are most useful for an individual prediction task.

We then illustrated RBP by predicting VORP (value over replacement player) for the rookie seasons for players who were drafted in 2022, 2023, and 2024. Our analysis revealed that RBP successively distinguished in advance which players would produce more favorable outcomes from those who would produce less favorable outcomes, and it did so more effectively than the draft in all cases. And we provided compelling evidence that, based on fit, we could distinguish in advance which predictions to trust and which to discard or treat with caution. We also highlighted the extreme transparency of RBP, which reveals precisely how each player informs an individual prediction and how each predictive variable contributes to the reliability of the prediction.

To conclude, we acknowledge that scouting information provides insights that would be unobtainable from any analytical system. However, we wish to emphasize that RBP produces valuable information that is unknowable from scouting as well as from other prediction techniques.

We would like to thank Miles Kee for valuable insights and computational assistance.

**Appendix: Convergence of Relevance to Other Prediction Methods**

Convergence to Linear Regression Analysis

The prediction equation corresponding to full sample linear regression equals:

$$\hat{y}_t = \bar{y} + \frac{1}{N-1}\sum_{i=1}^{N} r_{it}(y_i - \bar{y}) \quad \text{(A1)}$$

Expanding the expression for relevance gives:

$$\hat{y}_t = \bar{y} + (x_t - \bar{x})\frac{1}{N-1}\sum_{i=1}^{N}\Omega^{-1}(x_i - \bar{x})'(y_i - \bar{y}) \quad \text{(A2)}$$

To streamline the arithmetic, we recast this expression using matrix notation:

$$X_d = (X - 1_N\bar{x}) \quad \text{(A3)}$$

$$\hat{y}_t = \bar{y} - \bar{x}\beta + x_t\beta - (x_t - \bar{x})(X_d'X_d)^{-1}X_d'1_N\bar{y} \quad \text{(A4)}$$

Where:

$$\beta = (X_d'X_d)^{-1}X_d'Y \quad \text{(A5)}$$

Noting that $X_d'1_N$ equals a vector of zeros, because $X_d$ represents attribute deviations from their own respective averages, we get the familiar linear regression prediction formula:

$$\hat{y}_t = (\bar{y} - \bar{x}\beta) + x_t\beta \quad \text{(A6)}$$

$$\alpha = (\bar{y} - \bar{x}\beta) \quad \text{(A7)}$$

$$\hat{y}_t = \alpha + x_t\beta \quad \text{(A8)}$$

Relationship to Large Language Models

The key innovation that led to the success of large language models (LLMs) is the transformer, which is an information processing architecture based on attention mechanisms. Relevance is conceptually similar to attention and offers a novel interpretation of these models.

In the context of language processing, consider a sequence of words (or tokens) which is encoded as a vector, $x_i$. The goal is to transform each word into an enriched vector, $z_i$, with new dimensions, which represents a refined contextual meaning of the word within the passage.

As noted in Vaswani et al. (2017), attention in a transformer model is determined by a set of three transformation matrices: $W^Q$, $W^K$, and $W^V$, which compute what are commonly referred to as query, key, and value vectors from each word $x_i$. To highlight the link with RBP, we characterize this as follows:

$$q_t = x_t W^Q \tag{A9}$$

$$k_i = x_i W^K \tag{A10}$$

$$v_i = x_i W^V \tag{A11}$$

$$z_i = \sum_i softmax\left(\frac{q_t k_i'}{\sqrt{params}}\right) v_i \tag{A12}$$

We may intuitively think of $v_i$ as representing the learned unconditional meaning of each word in the passage. These values represent the dependent variable, and we want to predict the contextual meaning as a weighted average of $v_i$ for all words in the passage based on their relevance to $x_i$. We may express:

$$q_t k_i' = x_t W^Q W^K x_i' \quad \text{(A13)}$$

Equation A13 matches the definition of relevance in Equation 1 from earlier, if we assume $\bar{x} = 0$ and we have $W^Q W^K$ rather than the inverse covariance matrix to relate circumstances to each other. In other words, the learned matrices $W^Q W^K$ amount to a square matrix that is used to evaluate relevance. The letters used to characterize words are mostly arbitrary (compared to meaning), so learned mappings are necessary for language interpretation, whereas for meaningfully oriented data the inverse covariance matrix is well-motivated.

The softmax function serves as a censoring function that normalizes weights to sum to one, while also requiring them to be strictly positive. Thus, the use of softmax effectively censors observations to focus on the most relevant subset, similar to partial sample regression. There are many other complexities to transformers. We do not aim to provide a thorough accounting of how these models work. We merely wish to point out the striking similarity between the essence of the attention mechanism used in these models and the principles of RBP described in this article.

**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.” Working paper.

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

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

[1] Though we don't emphasize this advantage in this paper, please refer to Czasonis, Kritzman, and Turkington (2025a) for a thorough description of how RBP handles incomplete information.
[2] See, for example, Czasonis, Kritzman, and Turkington (2024) and Czasonis, Kritzman, and Turkington (2025c).
[3] VORP (value over replacement player) is an estimate of the points per 100 team possessions a player scores over a replacement player during the entire season assuming his teammates perform in line with the average of all NBA players. Replacement players are bench players and have a VORP of -2.
[4] See Czasonis, Kritzman, and Turkington (2023) for proof of this result.
[5] We also show in the Appendix that our definition of relevance aligns with the key breakthrough that enables large language models such as ChatGPT.
[6] See Czasonis, Kritzman, and Turkington (2023) for proof of this result.
[7] See Czasonis, Kritzman, and Turkington (2022b) for proof of this result.
[8] See Czasonis, Kritzman, and Turkington (2025a) for a thorough description of relevance-based importance.