PORTFOLIO MANAGEMENT RESEARCH
with. Intelligence

the journal of PORTFOLIO management

QUANTITATIVE TOOLS FOR ASSET MANAGEMENT

volume 49 number 9
SEPTEMBER 2023
pm-research.com

An Intuitive Guide to Relevance-Based Prediction

Megan Czasonis, Mark Kritzman, and David Turkington

Megan Czasonis

Megan Czasonis is a Managing Director and Head of Portfolio Management Research at State Street Associates. Her team collaborates with academic partners to develop new research on asset allocation, risk management, and quantitative investment strategy. The Portfolio Management Research team delivers this research to institutional investors through indicators, advisory projects, and thought leadership pieces. Megan has published numerous journal articles, is a co-author of "Prediction Revisited: The Importance of Observation," and her research received the 2022 Harry Markowitz Award for best paper in the Journal of Investment Management. Megan graduated Summa Cum Laude from Bentley University with a B.S. in Economics / Finance.

Mark Kritzman, CFA

Mark Kritzman, CFA is a Founding Partner and CEO of Windham Capital Management, LLC and the Chairman of Windham's investment committee. He is responsible for managing research activities and investment advisory services. He is also a Founding Partner of State Street Associates, and he teaches a graduate finance course at the Massachusetts Institute of Technology. Mark served as a Founding Director of the International Securities Exchange and as a Commissioner on the Group Insurance Commission of the Commonwealth of Massachusetts. He has also served on the Advisory Board of the Government Investment Corporation of Singapore (GIC) and the boards of the Institute for Quantitative Research in Finance, The Investment Fund for Foundations, and State Street Associates. He is currently a member of the Board of Directors of Protego Trust Company, the Advisory Board of the MIT Sloan Finance Group, the Board of Trustees of St. John's University, the Emerging Markets Review, the Journal of Alternative Investments, the Journal of Derivatives, the Journal of Investment Management, where he is Book Review Editor, and The Journal of Portfolio Management. He has written more than 100 articles for peer-reviewed journals and is the author or co-author of eight books including *Prediction Revisited: The Importance of Observation*, *Asset Allocation: From Theory to Practice and Beyond, Puzzles of Finance,* and *The Portable Financial Analyst*. Mark won Graham and Dodd scrolls in 1993, 2002, and 2010 the Research Prize from the Institute for Quantitative Investment Research in 1997, the Bernstein Fabozzi/Jacobs Levy Award 10 times, the Roger F. Murray Prize from the Q-Group in 2012, the Peter L. Bernstein Award in 2013 for Best Paper in an Institutional Investor Journal, and the Harry M. Markowitz Award for 2022. In 2004, Mark was elected a Batten Fellow at the Darden Graduate School of Business Administration, University of Virginia. Mark has a BS in economics from St. John's University, an MBA with distinction from New York University, and a CFA designation.

David Turkington

David Turkington is Senior Managing Director and Head of State Street Associates, State Street Global Markets' decades-long partnership with renowned academics that produces innovative research on markets and investment strategy. David is a frequent presenter at industry conferences, has published more than 30 research articles in a range of journals, and currently serves on the editorial board of the Journal of Alternative Investments. He is the co-author of three books including "Asset Allocation: From Theory to Practice and Beyond" and "Prediction Revisited: The Importance of Observation." His published research has received the 2010 Graham and Dodd Scroll Award, five Bernstein-Fabozzi/Jacobs-Levy Outstanding Article Awards, the 2013 Peter L. Bernstein Award for best paper in an Institutional Investor journal, the 2021 Roger F. Murray First Prize for outstanding research presented at the Q Group seminars, and the 2022 Harry Markowitz award for best paper in the Journal of Investment Management. David graduated summa cum laude from Tufts University with a BA in mathematics and quantitative economics, and he holds the CFA designation.

An Intuitive Guide to Relevance-Based Prediction

Megan Czasonis, Mark Kritzman, and David Turkington

Megan Czasonis
is a managing director and head of Portfolio Management Research at State Street Associates in Cambridge, MA.
mczasonis@statestreet.com

Mark Kritzman
is the chief executive officer of Windham Capital Management in Boston, MA, and is a senior lecturer at the MIT Sloan School of Management in Cambridge, MA.
kritzman@mit.edu

David Turkington
is senior managing director and head of State Street Associates in Cambridge, MA.
dturkington@statestreet.com

KEY FINDINGS

- *Relevance* measures the importance of an observation to a prediction. It is composed of similarity and informativeness, both of which are measured as Mahalanobis distances.
- *Fit* measures the reliability of a specific prediction and aggregates to R-squared across all prediction tasks.
- *Codependence* is the notion that the selection of predictive variables depends on the chosen observations and the choice of observations depends on the selected variables.

ABSTRACT

Relevance-based prediction is a new approach to data-driven forecasting that serves as a favorable alternative to both linear regression analysis and machine learning. It follows from two seminal scientific innovations: Prasanta Mahalanobis' distance measure and Claude Shannon's information theory. Relevance-based prediction rests on three key tenets: 1) *relevance*, which measures the importance of an observation to a prediction; 2) *fit*, which measures the reliability of each individual prediction task; and 3) *codependence*, which holds that the choice of observations and predictive variables should be determined jointly for each individual prediction task.

Relevance-based prediction is a new, theoretically justified approach for forming predictions from data. It follows an intuitive, yet statistically rigorous, process of relying on past experiences as a guide to future outcomes. In doing so, it addresses complexities that are beyond the capacity of linear regression in a way that is more transparent and adaptive than machine-learning algorithms.

By measuring the relevance of past experiences in a statistical way, we can form data-driven predictions that can be explained intuitively in terms of past events and their relevance to today. For investors, this transparency provides a distinct advantage over machine-learning algorithms, which are notoriously opaque and often fail to provide intuition. Moreover, by viewing predictions from the lens of past experiences, we can focus on a subset of the most relevant observations, which allows us to address asymmetric and conditional relationships that are beyond the capacity of linear regression. Relevance also empowers us to judge the reliability of an individual prediction and adapt our choice of observations and variables according to the prediction circumstances. This adaptability offers yet another advantage over existing methods.

Our purpose in this article is to offer a non-technical, intuitive guide to relevance-based prediction.[1] We start by emphasizing the importance of past

[1] For additional details on relevance-based prediction including mathematical formulas, refer to Czasonis, Kritzman, and Turkington (2022a, 2022b, 2023).

experiences for forming predictions. Then, we describe three key tenets of relevance-based prediction: *relevance*, which measures the importance of an observation to a prediction; *fit*, which measures the reliability of an individual prediction; and *codependence*, which holds that the choice of observations depends on the selected variables and the selection of variables depends on the chosen observations. We then compare relevance-based prediction with linear regression and machine-learning algorithms from the perspective of three guiding principles: flexibility, adaptability, and non-arbitrariness.

MOTIVATION

It is intuitive to predict by learning from past experiences. For example, suppose we want to predict how a recession will impact stock returns. It is natural to recall past recessions and consider how stocks performed in each of those experiences. Of course, we do not need to have lived through all these events. We can turn to history books or long datasets to learn what happened. If we want to predict what may happen to stocks during a future recession, we have past experiences to draw upon.

For example, the left side of Exhibit 1 shows one-year stock returns following the start of 17 US recessions since 1926. The average return is 3%, as indicated by the horizonal bar. It would not be unreasonable, albeit somewhat naive, to predict that this will be the stock return following a future recession. However, there is considerable dispersion in these historical outcomes, including some highly negative returns. Therefore, we should not be overly confident in this prediction.

Now, suppose we want to predict the stock return in the year after the end of a recession. This time we gather all the returns following the end of the recession windows and we get a different distribution, as shown in the right side of Exhibit 1. Now the average stock return is highly positive: 20%. Moreover, many of these past returns lie near the average. As a result, we may have greater confidence in this prediction than the first. We will return to this insight in our discussion of fit.

This process of relying on past experiences is intuitive and enables us to see how each observation contributes to a prediction. Relevance-based prediction follows a similar approach but refines the process of this simple illustration. It uses a rich set of data to determine relevant observations in a precise mathematical way. It recognizes that some past events are more important than others; thus, we should overweight some and underweight, or even ignore, others. It also enables us to measure the reliability of an individual prediction and adapt our approach based on the circumstances of the prediction task. The result is a prediction system that is more transparent, more adaptive to new circumstances, and less arbitrary than existing techniques.

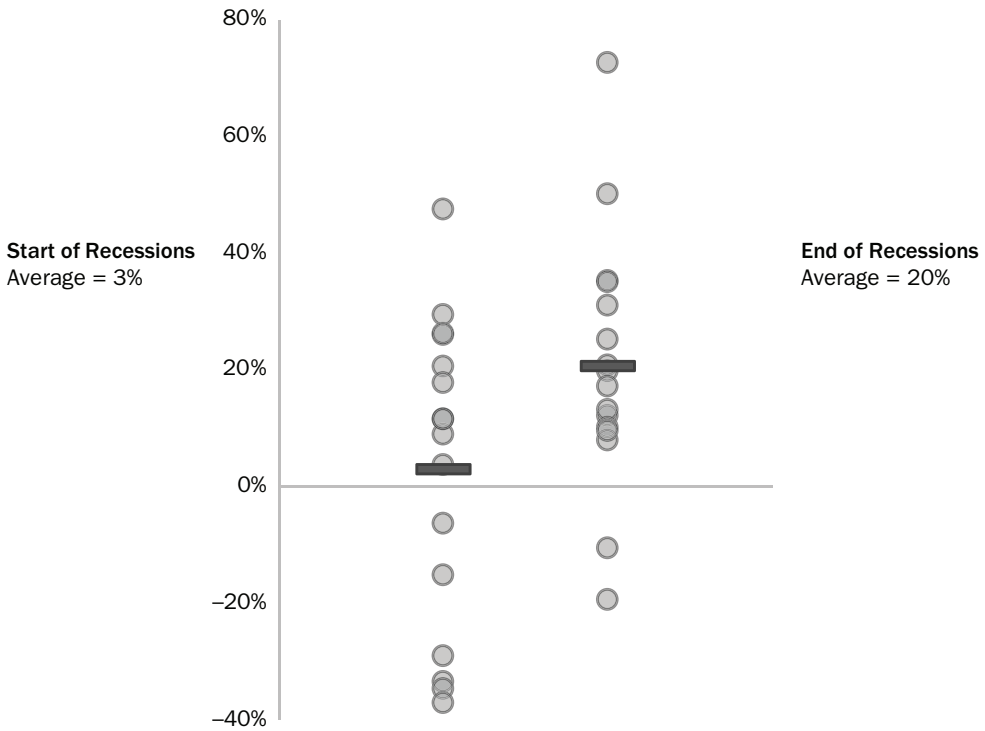
In fact, these are the three principles that guide relevance-based prediction. We propose that a prediction system should be the following:

1. *transparent*, which promotes intuition, facilitates interpretation, and inspires confidence;
2. *adaptive*, which means it responds to new circumstances;
3. *non-arbitrary*, by which we mean it is theoretically justified and mathematically unified.

We will reference these principles as we describe relevance-based prediction in the following sections. Later, we will use them to compare relevance-based prediction with other data analysis methods, namely, linear regression and machine learning.

EXHIBIT 1

One-Year Stock Returns Following Past US Recessions



NOTE: Horizontal bar indicates average return.

RELEVANCE

Let us start by defining relevance. Relevance is the importance of an observation to a prediction. It consists of two components: informativeness and similarity. Simply put, relevant observations are those that are different from average but are similar to today. Let us dig into these concepts further.

Informativeness is based on the notion that we should pay attention to unusual observations because they contain more information than typical ones. The link between unusualness and information is based on information theory, which was developed by Claude Shannon—one of the greatest geniuses in modern history—in the mid-20th century.[2] Information theory states that probability is inversely related to information. In other words, the more surprising an observation, the more information it contains. Interestingly, our emphasis on unusual observations challenges a maxim of the way classical statistics is often taught, which is that we should view unusual observations skeptically and seek to minimize their influence. However, the notion of overweighting unusual observations (assuming they are not errors, of course!) is theoretically grounded and intuitive. For example, suppose you are investigating the relationship between a particular factor and stock market performance. Would you feel more confident that you understand its impact on the market on a typical day when the factor barely moved or on a day when the factor moved dramatically and so did the market? Unusual observations contain more information.

The second component of relevance—similarity—is intuitive. It is based on the notion that we should pay attention to observations that resemble today. Investors

[2] For additional details, refer to Shannon (1948).

implicitly recognize this when they use rolling windows or overweight recent data points when running analyses. These common practices equate similarity with recency. However, they ignore the possibility that older periods may also bear semblance to the current environment. Our definition of similarity looks for common statistical patterns between past observations and today, regardless of where they fall along the timeline. This provides a richer set of experiences from which to form our predictions.

Together, informativeness and similarity form relevance. But how do we quantify these components? We rely on a powerful statistic called the Mahalanobis distance to measure them in a precise mathematical way. Named after Prasanta Mahalanobis—an esteemed statistician from the early 1900s—the Mahalanobis distance is a measure of multivariate distance between two data points.[3] It was originally conceived by Mahalanobis as part of an archeological study of human skulls. Mahalanobis devised the statistic to measure the similarity of a given skull—described by a set of attributes such as its length, the size of the nose, and so on—to the average attributes of skulls from a particular group. In developing this measure, Mahalanobis had two critical insights. First, he recognized that it is important to account for the typical deviation of a given attribute. For example, a one-centimeter difference in the size of a nose is more unusual than a one-centimeter difference in skull length. And second, he understood that it is important to account for the typical relationship between attributes. For example, if longer skulls tend to have larger noses, it would be unusual to find a long skull with a small nose. In mathematical terms, Mahalanobis realized that it was important to account for the variances of the attributes as well as their correlations.

We apply this same statistic to measure the informativeness and similarity of observations. In our case, we describe each observation by a set of circumstances related to the outcome we wish to predict. For example, these could be economic circumstances, such as growth and inflation, or market conditions, such as valuations and volatility. In fact, we can consider any set of circumstances that we think are important. In the language of data analysis, these are the X variables. Then, for each historical observation, we measure its informativeness as the Mahalanobis distance between its circumstances and the average circumstances across all observations. The greater its distance from average, the more unusual—and thus informative—the observation. We measure its similarity as the negative of its Mahalanobis distance from today's circumstances.[4] We use the negative distance because distance is a measure of dissimilarity and thus its negative indicates similarity. Finally, for each observation, we add these two components—informativeness and similarity—to determine its relevance. Relevant observations are those that are dissimilar from average (informativeness) and similar to today (similarity).

Prediction

Once we determine the relevance of all the observations in our sample, prediction follows naturally. We simply form our prediction as a relevance-weighted average of past outcomes for the unknown outcome we wish to predict. For example, these outcomes could be subsequent stock returns or GDP growth. In the language of data analysis, the past outcomes are the Y variable, though the X variables determine relevance.

[3] For additional details, refer to Mahalanobis (1927, 1936).

[4] More precisely, we measure similarity as half the negative Mahalanobis distance from current circumstances. We include a factor of one-half because distances between two data points are twice as large as distances from the average.

Similar to our simple illustration from earlier, this process is intuitive and enables us to see how each observation contributes to a prediction. Moreover, it has an important equivalence with linear regression. When we form our prediction as a relevance-weighted average of all historical outcomes, the resulting prediction is identical to that from a linear regression. This equivalence is important for two reasons. First, it means that relevance-based prediction is not arbitrary. It inherits the theoretical justifications of linear regression by virtue of their equivalence. Second, it reveals an interesting feature of linear regression. Linear regression assumes that what happened during relevant periods in the past will recur, which is reasonable. However, it also assumes that what happened during non-relevant periods in the past will recur, but in the opposite direction.

This is the essence of linearity; it expects opposite outcomes from opposite circumstances. However, this reasoning is occasionally, and perhaps often, flawed. For example, imagine asking your doctor how she thinks you will respond to a certain medication, and she explains that all of her patients who are different than you—different age, different sex, and the like—had horrible responses, but (don't worry!) yours should be the opposite. How confident would you be in her prediction? Or, returning to our recession example from earlier, it would be like studying stock returns from robust growth periods and predicting the opposite will occur during recessions. The reality is that the world is often complex and predicting the opposite of the least relevant experiences is not always a good idea. Intuitively, it makes more sense to ignore these observations, which relevance empowers us to do.

Partial Sample Regression

Forming predictions from a subset of relevant experiences is called partial sample regression. The idea is to censor the historical sample by removing non-relevant observations before forming our prediction as a weighted average of past outcomes. Each time we predict, we select a new subsample of observations that are relevant to the prevailing circumstances. This loosens the binds of linearity (at least across the full set of observations) and enables us to better predict outcomes that may rise from asymmetric or conditional relationships. In the presence of complexity, focusing on a subset of relevant observations often produces more reliable predictions.

FIT

Let us explore what we mean by reliability. Reliability is the quality of an individual prediction that we can assess in advance based on historical patterns. Importantly, this perspective is only possible when we understand how each past observation contributes to a prediction. To quantify reliability, we use a measure called *fit*.

Fit considers the strength of patterns in the data underlying a prediction. Recall that we form a prediction as a relevance-weighted average of past outcomes. Thus, each historical observation has a relevance weight and an outcome, the product of which is their contribution to the prediction. Fit quantifies the degree of alignment between these two components—relevance and outcomes—across the sample of observations underlying a prediction. The greater the alignment, the greater our confidence in the prediction.

To understand why, consider the following analogy. Suppose you ask six friends to predict the price of a bottle of Italian wine that you intend to purchase. Imagine that three of your friends know Italian wine well, and they all say this bottle will be expensive. Meanwhile, your other three friends who know little about wine think it will be cheap. The fact that your knowledgeable friends agree is comforting, and you

should be confident in their collective opinion. This is an example of good fit. Now, suppose you ask the same group of friends about an obscure French wine. The ones who know French wine disagree strongly: One thinks it is expensive and the other thinks it is cheap. This should make you uncertain about their prediction. This is an example of poor fit.

It is important to emphasize that fit is specific to an individual prediction. It acknowledges that it is harder to predict in some situations than others. In one extreme, we could have pure noise, with outcomes that are all over the place. In this case, prediction is a futile attempt, and fit alerts us to that. However, we could use the same approach and the same data in a different situation and find that outcomes and relevance align perfectly.

Mathematically, we measure fit as the squared correlation between relevance weights and outcomes. Here, there is another important mathematical equivalence. The R-squared of a linear regression model is exactly equal to the weighted average fit across all of its individual predictions. This equivalence is important because it means that our definition of fit is not arbitrary. Moreover, it underscores the point that R-squared is a summary statistic—an average of some good predictions and some bad ones. Fit provides greater transparency than R-squared because we can see how it changes for each prediction task. Moreover, because fit is determined from patterns in the data, we can estimate it before we even form a prediction. This paves the way for greater flexibility, allowing us to adapt our approach for each prediction task.

CODEPENDENCE

Each time we predict with relevance, we must make two key decisions: First, the extent to which we focus on a subset of relevant observations. And second, the set of variables we use to determine relevance. In fact, these two decisions are codependent. We should not choose observations independently of our variables, and we should not select variables independently of our observations. We should make these choices jointly.

Let us start with focus. The question is whether we should use every past observation in our prediction, as in linear regression, or instead focus on a subset of the most relevant observations. This decision entails a trade-off. As we narrow our focus, we might see stronger alignment among the data in the relevant subsample. However, this also introduces noise to the prediction because the subsample has fewer observations. Therefore, we should narrow our focus to the extent that the benefit of doing so (stronger subsample alignment) outweighs the cost (incremental noise). Fortunately, we can use fit to calibrate this decision. Fit recognizes the benefits and the costs of censoring data and quantifies them accordingly. Therefore, when forming a prediction, we can try a range of relevance thresholds for censoring the sample and select the one with the highest fit. Every time we predict, we can choose a different threshold.

Now let us turn to variable selection. To determine relevance, we must choose a collection of variables. This decision also entails a trade-off. Using a large number of variables can provide a richer picture of relevance. If a variable contributes more noise than information, however, it will harm our determination of relevance. Here, again, we can use fit to guide our decision. In this case, we can determine relevance using different groups of variables and select the group that maximizes fit. Every time we predict, we can choose a different set of variables that matter most to the prevailing environment.

These two decisions—focus and variable selection—are codependent. In order to select a subset of relevant observations, we need variables to determine their

relevance; and in order to select a group of variables, we need observations to determine if they are useful. Given this codependence, it is optimal to make these choices jointly. We refer to this process as CKT regression. For a given prediction task, we consider various combinations of relevance thresholds (for determining focus) and variables (for determining relevance), and we select the combination that maximizes fit. We repeat this process for each new prediction, adapting our choices in response to the new circumstances.

COMPARISON TO LINEAR REGRESSION AND MACHINE LEARNING

How does relevance-based prediction compare with other data analysis methods? Here we consider linear regression and machine learning, and we compare them with relevance-based prediction according to our guiding principles of transparency, adaptability, and non-arbitrariness.

Linear Regression

Let us start with linear regression, a staple of data analysis for the last two centuries. Linear regression analysis focuses on the selection of predictive variables (or X variables) and weights them based on an assumed linear relationship with outcomes (the Y variable). These weights, or betas, are determined by fitting a line through a scatter plot of observed values for the predictor variables and outcomes.[5] Then, the prediction for an unknown outcome is simply the beta-weighted average of the corresponding values for the X variables.

Linear regression's greatest strength is its theoretical underpinning. It is an elegant approach founded on rigorous principles. However, it is limited in a significant way. Linear regression fails in the presence of complexity such as an asymmetric relationship between predictive variables and outcomes. Moreover, although it offers transparency as to the importance of each predictive variable and a model's average reliability, it is opaque with respect to the importance of each observation and the reliability of an individual prediction.

To summarize:

- *Transparency*: Linear regression provides some transparency but is silent about how each observation informs a prediction and does not provide information about the quality of a specific prediction.
- *Adaptability*: Linear regression works well only if the influence of the predictive variables on the prediction is static across all observations. It fails in the presence of complex relationships between predictive variables and outcomes.
- *Non-arbitrariness*: Linear regression analysis is theoretically justified.

Machine Learning

Now let us turn to machine learning, which addresses complexities that are beyond the capacity of linear regression. For our purposes, it is helpful to classify machine-learning techniques into two categories: model-based and model-free algorithms.

[5] More precisely, it fits a line such that the sum of the squared distances of the observations from the line is minimized. Carl Friedrich Gauss, who originated this method of least squares circa 1795, proved that it gives a prediction whose expected variance from the truth is lower than any other linear and unbiased estimate.

Model-based algorithms are enhancements to linear regression analysis. Examples of model-based algorithms include lasso regression, tree-based algorithms, and neural networks. These techniques rely on an iterative process whereby you specify a decision rule, calibrate the rule, and then test the rule until you are satisfied with the results. Model-based algorithms are powerful in that they can be extraordinarily flexible in how they approach a wide range of complexities. After their final calibration, however, they are incapable of adapting to new circumstances. Moreover, model-based algorithms are notoriously opaque and rely on trial and error rather than theory.

Model-free algorithms form their predictions as weighted averages of prior outcomes. In this respect, they are similar to relevance-based prediction. Examples of model-free algorithms include near neighbors and Gaussian kernels. These models are adaptive in that they revise their weights with each new prediction task. However, the weights themselves lack a theoretical core.

To summarize:

- *Transparency*: The most powerful model-based algorithms are opaque and hard to interpret.
- *Adaptability*: Machine-learning algorithms do not adapt to new prediction circumstances.
- *Non-arbitrariness*: Machine learning is guided by empirical efficacy rather than by a core set of theoretical principles. They are, therefore, susceptible to algo mining.

Relevance-Based Prediction

This brings us to relevance-based prediction. Like linear regression, relevance-based prediction is not arbitrary. It is theoretically justified and mathematically unified with linear regression. Moreover, it improves upon linear regression in two key ways. First, it provides transparency into the importance of each observation as well as the reliability of an individual prediction. And second, it can address such complexities as asymmetric or conditional relationships, just as machine learning does. However, relevance-based prediction is more transparent, more adaptive to new circumstances, and less arbitrary than most machine-learning algorithms.

To summarize:

- *Transparency*: Relevance-based prediction shows precisely how the observations inform the prediction, and it quantifies the confidence we should assign to each unique prediction task.
- *Adaptability*: Relevance-based prediction identifies the optimal sample of observations and collection of variables simultaneously for each prediction task based on the prevailing circumstances of the prediction task.
- *Non-arbitrariness*: Relevance-based prediction is theoretically justified and mathematically unified with linear regression analysis.

CONCLUSION

Relevance-based prediction is a new, theoretically justified approach for forming predictions from data, which addresses complexities that are beyond the capacity of linear regression analysis in a way that is more transparent and more adaptive to new circumstances than machine-learning algorithms. It enables us to determine the relevance of past experiences based on their informativeness and similarity to current

circumstances, in a precise mathematical way. This observation-centric perspective paves the way for greater transparency and flexibility. It allows us to form predictions from a subset of relevant observations, which helps address complex relationships. It enables us to judge the reliability of an individual prediction based on historical patterns in the data. And it empowers us to adapt our choice of observations and variables for each prediction task.

ACKNOWLEDGMENTS

This material is for informational purposes only. The views expressed in this material are the views of the authors, are provided "as-is" at the time of first publication, are not intended for distribution to any person or entity in any jurisdiction where such distribution or use would be contrary to applicable law, and are not an offer or solicitation to buy or sell securities or any product. The views expressed do not necessarily represent the views of Windham Capital Management, State Street Global Markets, or State Street Corporation and its affiliates.

REFERENCES

Czasonis, M., M. Kritzman, and D. Turkington. 2022a. *Prediction Revisited: The Importance of Observation*. John S. Wiley & Sons.

——. 2022b. "Relevance." *Journal of Investment Management* 20 (1): 37–47.

——. 2023. "Relevance-Based Prediction: A Transparent and Adaptive Alternative to Machine Learning." *The Journal of Financial Data Science* 5 (1): 27–46.

Mahalanobis, P. C. 1927. "Analysis of Race-Mixture in Bengal." *Journal of the Asiatic Society of Bengal* 23: 301–333.

——, P. C. 1936. "On the Generalised Distance in Statistics." *Proceedings of the National Institute of Sciences of India* 2 (1): 49–55.

Shannon, C. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (July, October): 379–423, 623–656.