MIT Sloan School of Management

MIT Sloan School Working Paper 6417-21

Relevance

Megan Czasonis, Mark Kritzman, and David Turkington

March 10, 2021

RELEVANCE

Megan Czasonis
mczasonis@statestreet.com

Mark Kritzman
kritzman@mit.edu

David Turkington
dturkington@statestreet.com

THIS VERSION: MARCH 10, 2021

Abstract

The authors describe a new statistical concept called relevance from a conceptual and mathematical perspective, and based on their mathematical framework, they present a unified theory of relevance, regressions, and event studies. They also include numerical examples of how relevance is used to forecast.

RELEVANCE

Introduction

We all face the task of forecasting future outcomes of random variables. To do so, we must identify independent variables we believe are relevant to explaining variation in the dependent variable we would like to forecast. Then we need to collect observations of these independent variables. Although we think long and hard about which independent variables are relevant to our forecasts, we are typically less concerned about which observations of those variables are relevant.

Classical statistics tells us to collect as many observations as possible, because it is thought that a larger sample yields a more reliable forecast than a smaller sample. Yet we sometimes use rolling windows of observations rather than an expanding history, or we exponentially weight observations over time. These practices imply that we believe some observations are more relevant to our forecast than others, which is true. But it turns out that relevance is more complex than recentness, and that by accounting for the complexity of relevance, we may be able to produce better forecasts than by ignoring it or equating it to recentness.

In this article we proceed as follows. First, we describe the notion of relevance conceptually, including how it is used to improve forecasting. We then explain relevance mathematically, and we use this framework to present a unified theory of relevance, regressions, and event studies. Next, we give numerical examples to illustrate the application of relevance to different methods of forecasting. We conclude with a summary.

Relevance Conceptually

Relevance is a measure of the importance of an observation to a prediction. In regression analysis, relevance is determined by the independent variables, and the prediction pertains to a dependent variable. Relevance has two components, similarity and informativeness. This definition is hardly arbitrary. It follows from a mathematical equivalence we discuss later.

First, let us consider similarity. As noted, it is common to discriminate between more recent and less recent observations when compiling data samples for use in forecasting models. This practice of censoring or de-emphasizing older observations is often quite helpful, especially if the system that produces the observations undergoes structural change. The implicit assumption is that if recent data is more like current conditions it is more relevant and more reliable. But there are better ways to determine the relevance of the observations in a sample. In fact, human judgment and intuition provide an effective filter for relevance. People often look to history for experiences that are like current circumstances and use those similar experiences to provide guidance about how the future will unfold, irrespective of the chronological position of those similar experiences. Intuitively, it makes sense to consider similarity of past observations to current conditions as one component of relevance. This perspective, of course, does not necessarily exclude recent experiences as relevant.

Now let us consider the other component of relevance, informativeness. Classical statistics tells that if we do exclude observations, we should exclude those that are most extreme because they might reflect errors or arise from unusual circumstances that are unlikely to reappear. While it certainly makes sense to exclude incorrect data, we should not exclude or de-emphasize correct, outlying observations. To the contrary, we should emphasize them because unusual observations are more likely to be associated with consequential events, whereas common observations may arise purely from noise. Unusual observations are more informative. This view of informativeness is consistent with information theory, which posits that information is inversely related to probability.[1]

We define the informativeness of an observation as its dissimilarity from average. For relevance, it is important that we consider the informativeness of current conditions along with the informativeness of a historical observation. When we include both, along with similarity, the resulting measure of relevance has an average value of zero across all observations in any sample. It is natural to use this threshold of zero to distinguish between observations that are relevant (positive values) and not relevant (negative values). If instead we exclude the informativeness of the current observation, relevance could sum to an arbitrarily large positive or negative value, with the consequence that we would struggle to distinguish clearly between the relevant and non-relevant observations. By including the informativeness of the current observation, the meaning of relevance shifts from a relative quantity to an absolute quantity.

To summarize, the relevance of an observation is the sum of its similarity to current conditions, its dissimilarity from average conditions, and the dissimilarity of current conditions from average conditions.

Similarity and informativeness are multivariate concepts. When we speak of observations and conditions, we have in mind a multivariate description of circumstances, specifically a vector of values for a set of independent variables. When we measure the similarity of a past observation to the current observation, we would like to consider not only the similarity of the values of each variable in isolation, but also the similarity of their co-occurrence. And when we measure the informativeness of an observation, we would like to consider both the dissimilarity of the values of each variable from average as well as the dissimilarity of their co-occurrence from their average co-occurrence. Put plainly, we would like to consider how variables behave independently as well as how they interact with each other when measuring similarity to current conditions or dissimilarity from average conditions.

We use a statistic called the Mahalanobis distance to measure these features of data precisely.[2] Unlike the standard Euclidean distance, the Mahalanobis distance accounts for the variances and correlations of variables. All else equal, two observations are more distant (less similar) if the spread between their values is large compared to the typical variance of those values. And all else equal, two observations are more distant if the pattern of differences between their values diverges from the typical pattern of differences in values. The Mahalanobis distance neatly summarizes these effects in a single number.

Why should we care about relevance? We should care because it allows us to use observations more effectively in forecasts. To understand how relevance improves forecasting, we first need to understand how it is related to regression analysis.[3] The prediction from a linear regression equation is mathematically equivalent to a weighted average of the historical values of the dependent variable in which the weights are the relevance of the independent variables.

This equivalence reveals an intriguing feature of regression analysis. Owing to the symmetry of the observations around a fitted regression line, regression analysis places as much importance on non-relevant observations as it does on relevant observations. It just flips the sign of the effect of the non-relevant observation on the dependent variable. This feature of regression analysis invites a fundamental question about forecasting: Are non-relevant observations as useful in forming a prediction as relevant ones? In some cases, they may be, but not always, and perhaps not usually. Suppose, for example, we wish to forecast the economic outcomes of a recession. Should we place as much importance on past conditions of robust growth as on past recessions? This is an empirical question, but we suspect that intuition is often right to suggest that relevant observations are more useful to a forecast than non-relevant observations.

This insight about how regression analysis treats relevant and non-relevant observations leads to the key innovation we propose for forecasting. Researchers should consider a two-step approach to forecasting. First, create a subsample of relevant observations. And second, form the prediction as a relevance-weighted average of the past values of the dependent variable in the subsample. This two-step approach to forecasting is called partial sample regression.[4]

One might ask why we should not simply apply regression analysis to the subsample of relevant observations. Why do we instead take a weighted average of the past values of the dependent variable? The answer is that the weights preserve valuable information about relevance in the context of the full sample. If we were to apply regression analysis to the relevant subsample, it would consider some of the relevant observations as not relevant and interpret them opposite to the way they should be used to inform the prediction.

Perhaps at this point it would be useful to summarize our concept of relevance.

1. The relevance of an observation is determined by the independent variables for the purpose of forecasting a dependent variable. It equals the sum of an observation's similarity to current conditions, its informativeness, which is measured as its dissimilarity from average conditions, and the informativeness of current conditions.
2. By including similarity in our definition of relevance, we are simply following intuition, which often directs us to consider past events that are like current conditions to help us think about the path forward.
3. Observations that are dissimilar from their average values are more informative than observations that are like their average values, because unusual observations are more likely to have arisen from consequential events, whereas common observations may simply reflect noise in the data.
4. We include the informativeness of current conditions because by including it, the relevance of all the observations sums to zero, which establishes zero as a natural threshold for relevant and non-relevant observations.
5. To measure an observation's similarity to current conditions we should consider the isolated similarity of the variables' values to current values, as well as the similarity of their co-occurrence to the co-occurrence of current values. The same is true for how we

measure dissimilarity from the average values to determine informativeness. We should consider the values of the variables in isolation as well as how they interact with each other.

6. We use a statistic called the Mahalanobis distance to measure similarity and informativeness. The Mahalanobis distance considers variables independently as well as how they interact with each other. It also converts all values into common units.
7. The prediction from a linear regression model is mathematically equivalent to a relevance-weighted average of the past values of the dependent variable if it is averaged over the full sample.
8. This equivalence reveals that regression analysis places as much importance on non-relevant observations as it does on relevant observations, which is often counterproductive.
9. We should therefore consider forming our prediction as a relevance-weighted average of the dependent variable from a subsample of observations that have positive relevance.
10. We should not, however, apply regression analysis to a subset of relevant observations, because it will interpret some of the relevant observations in a way that is opposite to how they should inform the prediction.

Relevance Mathematically

In our conceptual discussion, we explained relevance within the context of a current observation and past observations. We now define it more generally between any pair of observations x_i and x_j, each of which is a row vector of values for a set of independent variables *X*. We define similarity and informativeness in terms of these two vectors and Ω^{-1}, the inverse of the full sample covariance matrix of *X*, as shown in Equations 1 and 2.

$$sim_{ij} = sim(x_i, x_j) = -\tfrac{1}{2}(x_i - x_j)\Omega^{-1}(x_i - x_j)' \tag{1}$$

$$info_i = info(x_i) = \tfrac{1}{2}(x_i - \bar{x})\Omega^{-1}(x_i - \bar{x})' \tag{2}$$

Similarity equals the Mahalanobis distance between x_i and x_j, in its squared form, multiplied by negative 1/2. It may be helpful to consider the purpose of each step in this calculation. The spread between the vectors measures the similarity of the values for each variable in isolation. Multiplying by the inverse of the covariance matrix converts the spreads for each variable into common units, effectively dividing each spread by the variance of the corresponding variable. It also captures the similarity of the co-occurrence of the variables compared to their typical patterns of co-occurrence. When we post multiply by the spreads between the vectors, we collapse the result into a single number. The negative sign converts the notion of distance into one of closeness (similarity). The factor of 1/2 offsets the double counting that occurs from the identical multiplication of x_i with x_j and x_j with x_i.

We measure Informativeness as the Mahalanobis distance between x_i and $\bar{x}$, the full sample mean of *X*, multiplied by 1/2. We multiply the square of the Mahalanobis distance by positive 1/2 because again we must offset the double counting that occurs from squaring $(x_i - \bar{x})$, but now we are interested in how dissimilar or distant the observations are from the average values, so we retain its positive value.

We define relevance as in Equation 3.

$$r_{ij} = r(x_i, x_j) = sim_{ij} + info_i + info_j \tag{3}$$

Recall from our earlier conceptual description that we include the informativeness of both x_i and x_j so that the relevance of all observations sums to zero. This result enables us to use a threshold of zero to separate relevant observations from non-relevant observations.

Relevance is independent of the object of our prediction, *Y*. In the absence of any information from the *X* variables, our best prediction $\hat{y}_t$ of an unknown y_t would be the simple average, $\bar{y}$. But the utility of relevance is that we may enhance that estimate by adding a weighted average of the historical deviations of *Y* from their average, where the weights are the relevance of each x_i to x_t.

$$\hat{y}_t = \bar{y} + \frac{1}{n-1}\sum_{i=1}^{n} r_{it}(y_i - \bar{y}) \tag{4}$$

A Unified Theory of Relevance, Regressions, and Event Studies

We now present a unified theory of relevance, regression analysis, and event studies. We proceed by illustrating the following facts. First, when we apply Equation 4 across the full sample of observations, $n = N$, we obtain the same forecast as a linear regression model. Second, when we apply the same procedure to a subsample of the most relevant observations, $n < N$, we obtain a valid partial sample regression forecast. Third, when we apply Equation 4 to a single observation, $n = 1$, which we choose for any reason, we end up with the outcome of a single event. And fourth, when we apply this procedure to a subsample of observations, $n < N$, which we choose for any reason, we obtain the results of a composite event study that is informed by the relevance of the observations.

Let us begin by demonstrating the equivalence between Equation 4 applied over the full sample and linear regression. First, we rearrange and consolidate the expression for relevance from Equation 3.

$$r_{it} = -\frac{1}{2}(x_i - x_t)\Omega^{-1}(x_i - x_t)' + \frac{1}{2}(x_i - \bar{x})\Omega^{-1}(x_i - \bar{x})' + \frac{1}{2}(x_t - \bar{x})\Omega^{-1}(x_t - \bar{x})' \quad (5)$$

$$r_{it} = x_t\Omega^{-1}x_i' - x_t\Omega^{-1}\bar{x}' - \bar{x}\Omega^{-1}x_i' + \bar{x}\Omega^{-1}\bar{x}' \quad (6)$$

$$r_{it} = (x_t - \bar{x})\Omega^{-1}(x_i - \bar{x})' \quad (7)$$

We substitute Equation 7 into Equation 5.

$$\hat{y}_t - \bar{y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_t - \bar{x})\Omega^{-1}(x_i - \bar{x})'(y_i - \bar{y}) \tag{8}$$

Equation 8 predicts the value of *Y* above its average based on observations of *X* above its average and *Y* above its average. The covariance matrix, by definition, is also a function of *X* above its average. Therefore, without loss of generality we may rewrite the prediction formula under the assumption that *X* and *Y* have means of zero. We pull $x_t\Omega^{-1}$ out of the sum because they do not depend on i.

$$\hat{y}_t = x_t\Omega^{-1}\frac{1}{n-1}\sum_{i=1}^{n} x_i' y_i \tag{9}$$

Using matrix notation whereby X contains all n observations in rows and k variables in columns, and Y contains all n observations in rows with one column and noting that $\Omega^{-1} = (n-1)(X'X)^{-1}$, we obtain the standard formula for a linear regression prediction.

$$\hat{y}_t = x_t(X'X)^{-1}X'Y \tag{10}$$

$$\hat{y}_t = x_t\beta' \tag{11}$$

Linear regression, and its relevance-weighted equivalent, do not discriminate between highly relevant and highly non-relevant observations other than flipping the sign of their predictive contribution. In cases where relevant observations are more reliable than non-relevant ones, it may be better to remove the non-relevant observations and apply Equation 4 to a relevant subsample of the data, $n < N$. Because we now estimate $\bar{y}$ on the subsample, and because $(y_i - \bar{y})$ sums to zero over the subsample, the expected value of our partial sample regression forecast remains properly centered on the subsample mean, even though the relevance weights are all positive.

Now consider an event study intended to give a prediction of the path of a chosen variable following an event that just occurred or is anticipated to occur. As a simple approach, we might identify a single past observation, $n = 1$, based on judgment, intuition, or exogenous variables, and record the outcome of *Y* at a range of time intervals after the event. The single observation we choose could be the most relevant one, but it need not be. In either case, we may consider each time interval observation of *Y* around the historical event as an application of Equation 4 with one data point. When $n = 1$, our prediction for *Y* converges to its actual occurrence following the event.

An event study with multiple events, $n < N$, is potentially more interesting and more statistically robust. To conduct a traditional composite event study, we identify a sample of events and align their chronological position to $t = 0$. We then observe the value of a chosen variable *Y* at various times following the event, $t + 1$, $t + 2$, $t + 3$, … , and we compute the arithmetic mean of these post-event observations. We interpret these post-event means as predictions for the path *Y* will take from a recent or anticipated event.

By selecting the events, we effectively censor non-relevant observations just as we do when we create a subsample of relevant observations from which to form our prediction in a partial sample regression, but we are using criteria other than the relevance of *X*. Now suppose that rather than predicting the path forward as the arithmetic mean of the observed paths, we weight the observed paths by their statistically determined relevance. This would be the same as weighting the observations by their relevance in partial sample regression, after removing non-events from consideration. Therefore, a relevance-weighted event study is equivalent to partial sample regression, with the exception that non-relevant observations are censored based on identification as non-events as opposed to the statistical relevance of *X*.

To summarize our unified theory:

1. The prediction from a linear regression equation is mathematically equivalent to a weighted average of past values of the dependent variable in which the weights are the relevance of the independent variables.
2. This equivalence allows one to form a relevance-weighted prediction of the dependent variable by using only a subsample of relevant observations. This approach is called partial sample regression.
3. Like partial sample regression, an event study separates relevant observations from non-relevant observations, but it does so by identification rather than mathematically.
4. As an alternative to predicting the path from a recent or current event as an arithmetic mean of past paths, one could use a relevance-weighted average of past paths to form a prediction. This approach would be equivalent to partial sample regression in which the

relevant subsample is determined by a separate identification process rather than statistical relevance.

5. Hence, the equivalence of relevance, regressions, and event studies.

Empirical Illustrations

Next, we illustrate the application of relevance to three forecasting methods: time series regression, cross-sectional regression, and event studies. These examples are intended to demonstrate how to incorporate relevance into each method and the intuitive appeal in doing so. As such, the models are admittedly simple, and the results are not intended as robust backtests of their efficacy.

Times Series Regression

First, we employ partial sample regression in a time series context to predict the winner of U.S. presidential elections. We illustrate our approach with the 2008 and 2016 elections, which provide an interesting comparison given their opposing outcomes. To generate our forecasts, we use a historical sample of presidential elections since 1876 and specify our model as follows:[5]

Dependent variable (Y):

- Percentage of electoral votes for the Democratic candidate.

Independent variables (X):

- Political variables:
  - Party affiliation of incumbent president (0 or 1)
  - Is the incumbent running for another term? (0 or 1)
  - Senate – Majority party (0 or 1)
  - Senate – Percentage of seats held by Democrats
  - House – Majority party (0 or 1)
  - House – Percentage of seats held by Democrats
- Geopolitical variable:
  - Was the U.S. at war during the election year? (0 or 1)
- Economic variables:
  - Was the U.S. in a recession during the election year? (0 or 1)
  - Trailing four-year economic growth, measured as percentage change in GDP
  - Trailing four-year change in debt, measured as change in Debt-to-GDP
  - Trailing four-year US stock return

We apply partial sample regression to forecast the 2008 and 2016 U.S. presidential election outcomes based on a subset of relevant historical elections. Relevant elections are those with positive relevance with respect to the election of interest, based on the independent variables. The predictions are out-of-sample, based only on data available as of July 31st of the election year and accounting for point-in-time economic data.

Exhibit 1 reports the model's predictions and actual outcomes for the percentage of electoral votes won by the Democratic candidate in 2008 and 2016. For comparison, we also include linear regression predictions based on the full sample of historical elections. Partial sample regression correctly predicted the presidential victor in both sample elections. Notably, it correctly predicted the 2016 outcome and linear regression did not. This illustrates the value in censoring non-relevant observations when generating forecasts.

Exhibit 1: Predictions and Realizations for the 2008 and 2016 U.S. Presidential Elections
(Percentage of electoral votes for Democratic candidate)

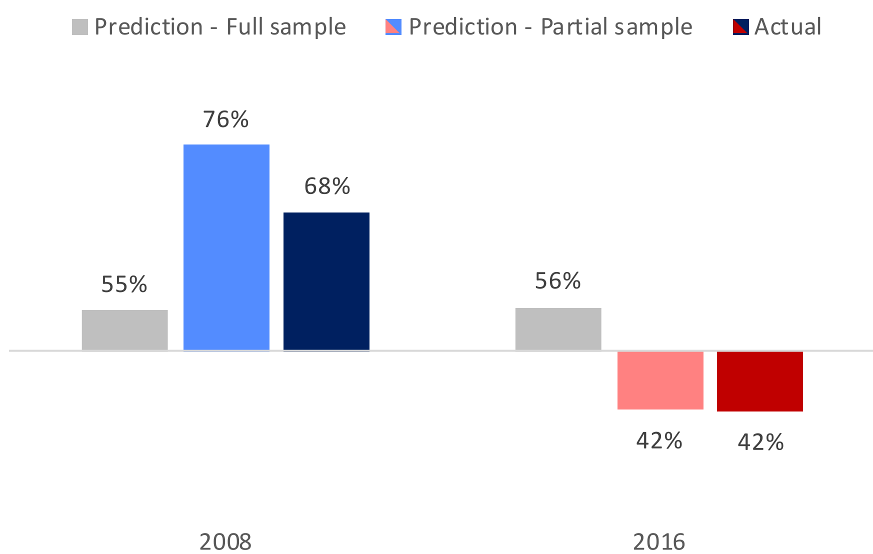
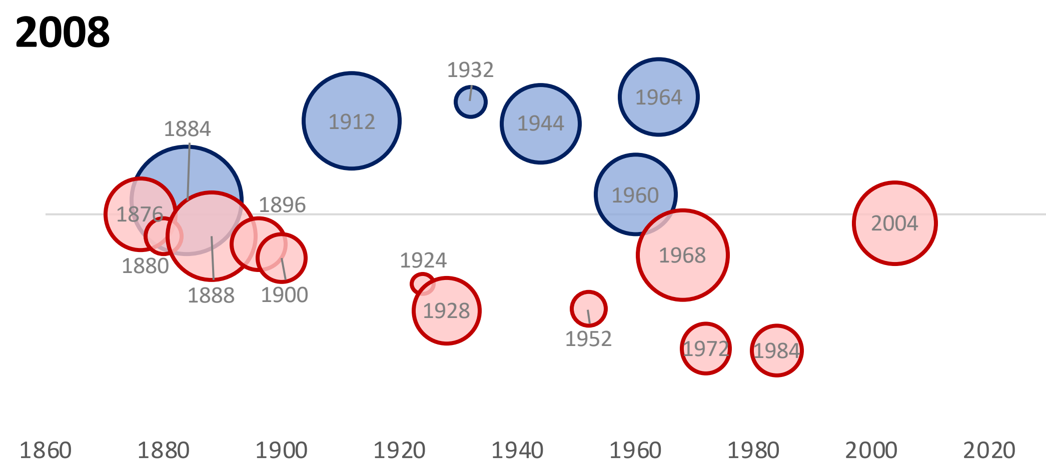


Exhibit 2 details the subset of relevant elections underlying the partial sample predictions. The height of each circle equals the percentage of electoral votes for the Democratic candidate (with color indicating the winner), and the area of each circle is proportional to the relevance of that observation.

Exhibit 2: Statistically Relevant Prior Elections and their Outcomes



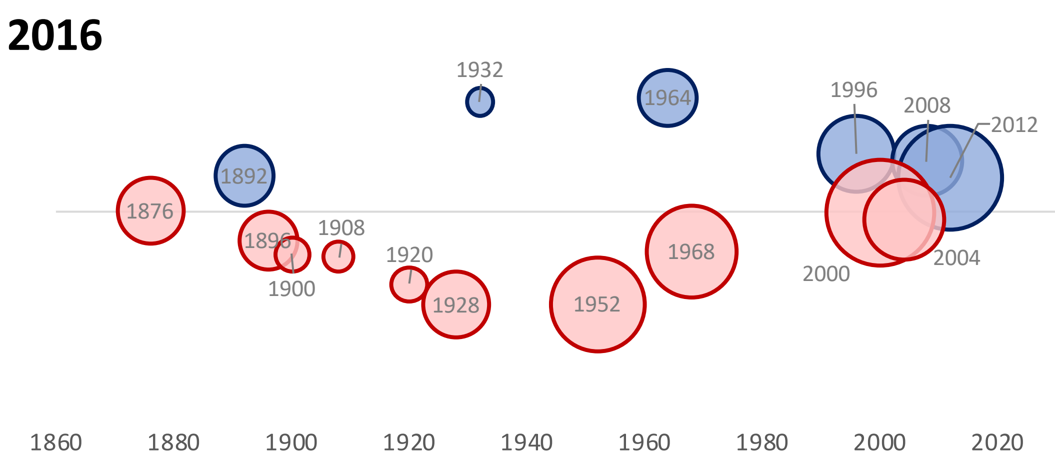


Exhibit 2 provides an interesting contrast between the two election forecasts. 2008 was generally reminiscent of older elections while 2016's forecast relied on more modern elections. For example, the two most relevant elections to 2008 were 1884 and 1912, while 2000 and 2012 were most relevant to 2016. It is also interesting to note differences in the dispersion of election outcomes between the two subsamples. For example, though the model correctly predicted a Democratic victory for 2008, only six of the 18 relevant historical elections had

Democratic victors. This suggests that the relative relevance of those six elections was enough to tilt the prediction in their direction. In contrast, 10 of the 16 relevant elections to 2016 were Republican victories, in line with the model's prediction.

These observations illustrate the value in viewing predictions through the lens of relevant observations. It offers intuition by comparing the predictors from the relevant observations to today. It instills confidence by showing the dispersion in outcomes across relevant observation. And, it yields unexpected insights by highlighting statistical adjacency over chronological adjacency.

Cross-Sectional Regression

Next, we apply relevance in a cross-sectional context to identify firms that are, collectively, in similar circumstances to a company of interest. In turn, an investor could predict various outcomes for the firm of interest, such as earnings announcements or stock price moves, based on this subset of relevant firms.

We illustrate our approach using S&P 500 constituents as of December 2019 and focus on Alphabet and Delta Airlines as our firms of interest. These companies provide an interesting comparison given their different sector classifications, fundamental attributes, and performance in the wake of the COVID pandemic. To identify a subset of comparable firms for Alphabet and Delta Airlines, respectively, we measure the relevance of S&P 500 constituents with respect to each company, based on the following firm attributes:

Independent variables (X):[6]

- Size (Log of market capitalization as percentage of the S&P 500)
- Value (Book-to-Price)
- Earnings yield (Earnings-to-Price)
- Momentum (Log of one plus the trailing 12-month price return)

Exhibit 3 summarizes the relevant firms for Alphabet (top panel) and Delta Airlines (bottom panel) as of December 2019. Relevant firms are those with positive relevance to the firm of interest, based on the variables described previously. The left tables report the fraction of firms within each sector that are relevant to the given company. The right tables show the top 10 most relevant firms to each.

Exhibit 3: Statistically Relevant Firms as of December 2019

Alphabet (Communication Services)

Fraction of relevant firms within sectors

Sector	Relevant firms
Energy	62%
Communication Services	57%
Health Care	54%
Consumer Staples	53%
Utilities	48%
Information Technology	45%
Financials	40%
Industrials	38%
Materials	33%
Real Estate	31%
Consumer Discretionary	30%

Top 10 most relevant firms

Name	Relevance	Sector
Microsoft	12.8	Information Technology
Facebook	10.8	Communication Services
Johnson & Johnson	10.1	Health Care
JP Morgan Chase & Co.	10.0	Financials
Exxon Mobile	9.7	Energy
Walmart	9.1	Consumer Staples
Visa	9.0	Information Technology
United Health Group	8.8	Health Care
AT&T	8.7	Communication Services
Chevron	8.6	Energy

Delta Airlines (Industrials)

Fraction of relevant firms within sectors

Sector	Relevant firms
Financials	81%
Communication Services	68%
Consumer Discretionary	56%
Utilities	52%
Consumer Staples	47%
Industrials	43%
Energy	42%
Health Care	42%
Information Technology	41%
Materials	33%
Real Estate	17%

Top 10 most relevant firms

Name	Relevance	Sector
NRG Energy	10.1	Utilities
Vornado Realty	5.8	Real Estate
Unum Group	3.8	Financials
Viacom B	3.5	Communication Services
Johnson Controls Intl.	2.9	Industrials
Macy's	2.9	Consumer Discretionary
American Airlines	2.9	Industrials
Viacom CBS	2.8	Communication Services
General Motors	2.8	Consumer Discretionary
United Airlines	2.8	Industrials

Exhibit 3 yields interesting insights. First, consider the fraction of relevant firms within each sector. Alphabet's sector classification, Communication Services, ranks high on its list (second, behind Energy). More than half, 57%, of the firms in its sector are relevant. Compare that to Delta Airlines, where only 43% of the firms in its sector, Industrials, are relevant. In fact, five sectors rank higher in terms of the fraction of firms that are relevant. Notably, 81% of financial firms are relevant to Delta. Though in practice, further analysis should be done to validate these comparisons, this simple illustration highlights the benefit of determining relevance statistically. Human judgment may fail to identify firms that are statistically adjacent to a given company based a collection of attributes, though they may differ in terms of sector classification or based on a single attribute in isolation.

Nonetheless, there are intuitive comparisons as well. For example, Microsoft and Facebook are the top two most relevant firms to Alphabet. In the case of Delta Airlines, American Airlines and United Airlines rank among its top 10 most relevant firms.

Our purpose here is not to form specific predictions based on these subsets of relevant firms, but rather to illustrate the distinction between relevance determined statistically and relevance determined judgmentally. However, one might imagine using this type of analysis to project earnings for companies that have not yet reported, or earnings for private companies based on public firms in similar circumstances.

Event Study

Finally, we employ relevance in an event study framework to forecast interest rate paths following shifts in monetary policy. To illustrate our approach, we run two event studies. In the first, we predict the path of U.S. interest rates following January 2001, the start of an expansionary monetary policy regime. In the second, we predict the path of interest rates following June 2004, the start of a contractionary regime. To generate our forecasts, we collect monthly economic data beginning in September 1982 and structure our event studies as follows:

Events:[7]

- Event Study A (January 2001): 11 historical months corresponding to the start of expansionary monetary policy regimes
- Event Study B (June 2004): 11 historical months corresponding to the start of contractionary monetary policy regimes

Dependent variable (Y):[8]

- Subsequent 24-month path of the target federal funds rate

Independent variables (X):[9]

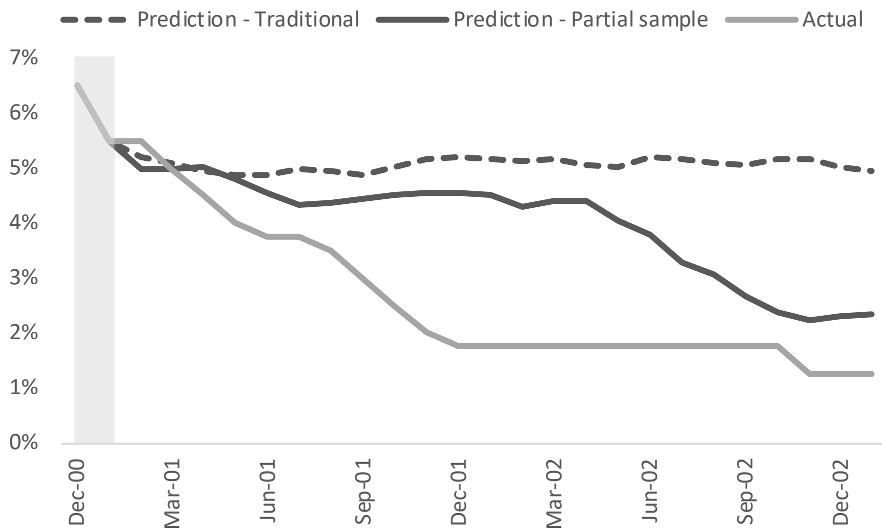
- Growth (12-month percentage change in industrial production)
- Inflation (12-month percentage change in CPI)
- Level of interest rates (Effective federal funds rate)
- Change in interest rates (12-month change in effective federal funds rate)

For each event study, we forecast the path of future interest rates by relevance-weighting observed paths following relevant prior events. In other words, we censor our historical sample to exclude both non-events and non-relevant events (those with negative relevance to the event of interest). Importantly, relevance here is measured in the context of the full sample of monthly observations. It is not measured within a subsample of pre-defined events for reasons described earlier.

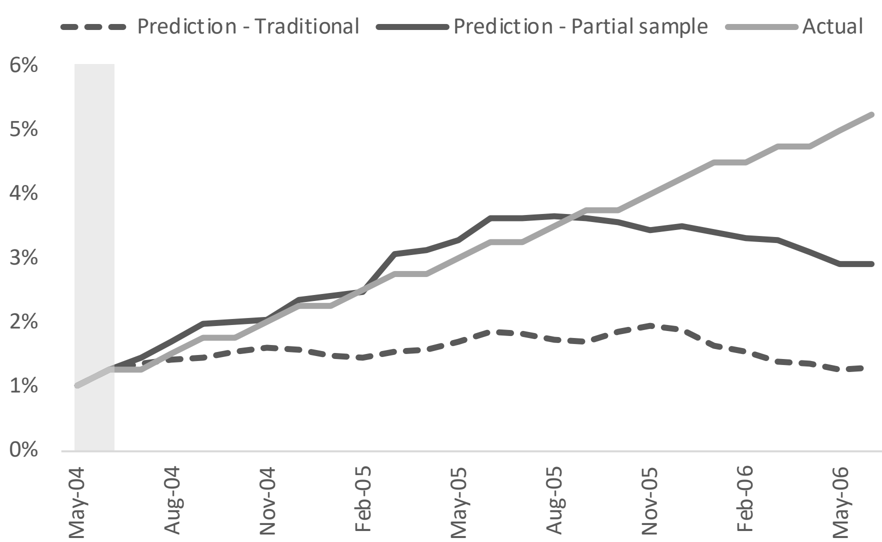
Exhibit 4 plots the predicted and realized interest rate paths for the two event studies. For comparison, we also include the forecasted paths from a traditional approach that equally weights observed paths across all historical events. It is interesting to note the difference between the traditional and partial sample forecasts. In both event studies, the traditional paths are highly muted compared to the partial sample-based paths. This illustrates the value in considering relevance, even across observations that are innately related, as in the case of pre-defined events. By focusing on a subset of events that are statistically adjacent to the one of interest and weighting them according to relevance, we generate more meaningful predictions that align more closely with actual outcomes.

Exhibit 4: Predicted and Realized Interest Rate Paths following Monetary Policy Events

Event Study A: January 2001



Event Study B: June 2004



Summary

Although most of us think long and hard about which variables to use in our forecasts, we typically tend not to think as much about which observations of those variables to include. To the extent we do consider observations, we are often inclined to place greater emphasis on more recent observations than more distant observations. However, when we think intuitively about how to forecast into the future from present conditions, we often look to past episodes in history that are like present conditions. This intuition is sound and helpful. Observations that are like current conditions are more relevant to a forecast than dissimilar observations. But not all observations that are equally like current conditions are equally relevant. Observations that are unusual are more relevant than common observations, because unusual observations are more likely to be associated with consequential events, whereas common observations are more likely to reflect noise. Thus, unusual observations are more informative.

The relevance of an observation is determined by its similarity to current conditions, its dissimilarity from average conditions (which captures its informativeness), and the informativeness of current conditions. We include the informativeness of current conditions to facilitate a natural interpretation of relevance in absolute terms. By including it, the relevance of all observations sums to zero, which enables us to use a threshold of zero to separate relevant observations from non-relevant observations.

When we measure relevance, not only must we measure the similarity of variable values to their current values or their dissimilarity from average values in isolation. We must also consider the similarity or dissimilarity of their co-occurrence. We therefore use a statistic called the Mahalanobis distance to measure similarity and informativeness. This statistic has two valuable features: it considers the interaction of the variables, and it converts their values into common units.

Our conception of relevance is not arbitrary. The prediction from a linear regression equation is mathematically equivalent to a weighted average of the past values of the dependent variable in which the weights are the relevance of the independent variables. This equivalence reveals a key insight about regression analysis, which is that owing to the symmetry of observations around a fitted regression line, regression analysis places as much importance on non-relevant observations as it does on relevant observations; It just flips the sign of the effect of the non-relevant observation on the dependent variable.

This insight about regression analysis invites a fundamental question. Is it possible to produce a better forecast from a subsample of relevant observations than from the full sample? The answer, of course, can only be determined empirically, but it is not hard to imagine settings in which our intuition would rightly suggest that we exclude non-relevant observations.

We should therefore consider a two-step approach to forecasting. First create a subsample of relevant observations. Then, form the forecast by taking a relevance-weighted average of the observations from the relevant subsample. This two-step approach to forecasting unifies relevance, regression analysis, and event studies.

Notes

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

Chow, G., E. Jacquier, K. Lowry, and M. Kritzman. 1999. "Optimal Portfolios in Good Times and Bad." *Financial Analysts Journal*, vol. 55, no. 3 (May/June): 65–73.

Czasonis, M., M. Kritzman and D. Turkington. 2020. "Addition by Subtraction: A Better Way to Predict Factor Returns (And Everything Else)." *The Journal of Portfolio Management*, vol. 46, no. 8 (September).

Czasonis, M., M. Kritzman and D. Turkington. 2020. "The Past as Prologue: A New Approach to Forecasting." Working Paper.

Czasonis, M., M. Kritzman, B. Pamir and D. Turkington. 2020. "Enhanced Scenario Analysis." *The Journal of Portfolio Management*, 46 (4): 69-79 (March).

Kritzman, M. and Y. Li. 2010. "Skulls, Financial Turbulence, and Risk Management." *Financial Analysts Journal*, vol. 66, no. 5 (September/October): 1-10.

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

Mahalanobis, P. C. 1936. "On the Generalised Distance in Statistics." *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1: 49–55.

Shannon, C. 1948. "A mathematical Theory of Communication." *The Bell System Technical Journal*, vol. 27, pp 379-423, 623-656, July, October.

[1] See, for example, Shannon (1948).

[2] The Mahalanobis distance was introduced by an Indian statistician in 1927 and modified in 1936 to analyze resemblances in human skulls among castes in India. Mahalanobis compared a set of measurements for a chosen skull to the average of those measurements across skulls from two separate castes. One set of skulls was collected from a graveyard, and the other set was collected from a distant battlefield. He also compared the co-occurrence of those measurements for a chosen skull to their covariation within the caste. He summarized these comparisons in a single number which he used to place a given skull in one caste or the other.

[3] When we use the terms regression or regression analysis, we have in mind ordinary least squares linear regression analysis.

[4] See Czasonis, Kritzman, and Turkington (2020a) for a thorough discussion of partial sample regression.

[5] See Czasonis, Kritzman, and Turkington (2020b) for a more details and out-of-sample predictions across all six U.S. presidential elections since 2000.

[6] We use the following stock-level data from WorldScope: Market capitalization (item 08001), Price/Book Value Ratio - Close (item 09304), and Price/Earnings Ratio – Close (item 09104). All attributes are as of Q4 2019. Our universe consists of 443 stocks with complete information.

[7] We define the start of an expansionary (contractionary) regime as the first month in a series of rate cuts (hikes) in the target federal funds rate.

[8] We use the Target Federal Funds Rate (DFEDTAR) from the Federal Reserve Bank of St. Louis' FRED database.

[9] We obtain the following data from the Federal Reserve Bank of St. Louis' FRED database: Industrial Production (INDPRO), CPI-U NSA (CPIAUCNS), and the Effective Federal Funds Rate (DFF). We convert the effective federal funds rate to a monthly series by averaging its daily values over each calendar month. We lag all data by one month (such that it corresponds to the month prior to each event).