PORTFOLIO
MANAGEMENT
RESEARCH

the journal of
PORTFOLIO
management

volume 46
number 8

SEPTEMBER
2020

jpm.pm-research.com

Addition by Subtraction: *A Better Way to Forecast Factor Returns (and Everything Else)*

Megan Czasonis, Mark Kritzman, and David Turkington

MEGAN CZASONIS

Megan Czasonis is a Managing Director for the Portfolio and Risk Research team at State Street Associates. The Portfolio and Risk Research team collaborates with academic partners to develop new research on asset allocation, risk management, and investment strategy. The team delivers this research to institutional investors through indicators, advisory projects, and thought leadership pieces. Megan has co-authored various journal articles and works closely with institutional investors to develop customized solutions based on this research. Megan graduated Summa Cum Laude from Bentley University with a B.S. in Economics / Finance.

MARK KRITZMAN

Mark Kritzman is a Founding Partner and CEO of Windham Capital Management, a Founding Partner of State Street Associates, and teaches a graduate finance course at the Massachusetts Institute of Technology. Mark served as a Founding Director of the International Securities Exchange and has served on several Boards, including the Institute for Quantitative Research in Finance, The Investment Fund for Foundations, and State Street Associates. He has written numerous articles for academic and professional journals and is the author / co-author of seven books including *A Practitioner's Guide to Asset Allocation*, *Puzzles of Finance*, and *The Portable Financial Analyst*. Mark won Graham and Dodd Scrolls in 1993 and 2002, the Research Prize from the Institute for Quantitative Investment Research in 1997, the Bernstein Fabozzi/Jacobs Levy Award nine times, the Roger F. Murray Prize from the Q-Group in 2012, and the Peter L. Bernstein Award in 2013 for Best Paper in an Institutional Investor Journal. 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 Portfolio and Risk Research at State Street Associates. His team is responsible for research and advisory spanning asset allocation, risk management and quantitative investment strategy. Mr. Turkington is a frequent presenter at industry conferences, has published research articles in a range of journals, and led the development of State Street's systemic risk and turbulence indicators. He is also co-author of the book *A Practitioner's Guide to Asset Allocation*. His research has received the 2013 Peter L. Bernstein Award, four Bernstein-Fabozzi/Jacobs-Levy Outstanding Article Awards, and the 2010 Graham and Dodd Scroll Award. Mr. Turkington graduated summa cum laude from Tufts University with a BA in mathematics and quantitative economics, and he holds the CFA designation.

Addition by Subtraction: *A Better Way to Forecast Factor Returns (and Everything Else)*

Megan Czasonis, Mark Kritzman, and David Turkington

Megan Czasonis
is a managing director at State Street Associates in Cambridge, MA.
mczasonis@statestreet.com

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

David Turkington
is a senior managing director at State Street Associates in Cambridge, MA.
dturkington@statestreet.com

KEY FINDINGS

- 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 observations.
- Relevance within this context is defined as the sum of statistical similarity and informativeness, both of which are defined as Mahalanobis distances.
- Together, these features allow researchers to censor less relevant observations and derive more reliable predictions of the dependent variable.

ABSTRACT: *Financial analysts assume that the reliability of predictions derived from regression analysis improves with sample size. This is thought to be true because larger samples tend to produce less noisy results than smaller samples. But this is not always the case. Some observations are more relevant than others, and often one can obtain more reliable predictions by censoring observations that are not sufficiently relevant. The authors introduce a methodology for identifying relevant observations by recasting the prediction of a regression equation as a weighted average of the historical values of the dependent variable, in which the weights are the relevance of the independent variables. This equivalence allows them to use a subset of more relevant observations to forecast the dependent variable. The authors apply their methodology to forecast factor returns from economic variables.*

TOPICS: ***Portfolio management/multi-asset allocation, risk management, quantitative methods****

Financial analysts regularly use regression analysis to make predictions, and they typically assume that the quality of their predictions improves with sample size. This assumption is not always valid; quality instead depends on the extent to which the observations for the independent variable are equally relevant. It may be the case that a strong relationship exists within a subset of the observations but is concealed by observations that are not relevant. By censoring irrelevant observations, one might uncover these relationships.

We begin by introducing the notion of relevance as the sum of multivariate similarity and informativeness. We next show that the predicted value from a regression equation can be expressed as a weighted average of the prior values of the dependent variable, in which the weights equal the relevance of the prior values for the independent variables. We next discuss the merits of predicting the dependent variable from a subsample in which the irrelevant observations have been censored. Finally, we apply this methodology to forecast factor returns from economic variables. Our results show that by forecasting factor returns in this manner, we can produce more reliable predictions than by applying regression analysis to the full sample of observations.

RELEVANCE

We define *relevance* as the sum of multivariate similarity and informativeness. We start by using the Mahalanobis distance to measure the multivariate distance between two observations, as shown in Equation 1.[1]

$$d(x_i, x_t) = (x_i - x_t)\Omega^{-1}(x_i - x_t)' \tag{1}$$

Here, x_t is a vector of the current values of the independent variables, x_i is a vector of the prior values of the independent variables, the symbol $'$ indicates matrix transpose, and Ω^{-1} is the inverse covariance matrix of X where X comprises all the vectors of the independent variables.

The Mahalanobis distance takes into account not only how independently similar the components of the x_is are to those of the x_ts, but also the similarity of their co-occurrence to the co-occurrence of the x_ts. Multivariate similarity is simply the opposite (negative) of the multivariate distance between x_i and x_t:

$$Similarity(x_i, x_t) = -(x_i - x_t)\Omega^{-1}(x_i - x_t)' \tag{2}$$

All else equal, prior observations for the independent variables that are similar to current observations are more relevant than prior observations that are less similar.

However, not all measurements of similarity are alike. Observations that are close to their historical averages may be driven more by noise than by events. These ordinary occurrences are therefore less relevant. Observations that are distant from their historical averages are unusual and therefore more likely to be driven by events; these event-driven observations are potentially more informative.[2] Given this intuition, we define the informativeness of a prior observation x_i as its multivariate distance from its average value, $\bar{x}$:

$$Informativeness(x_i) = (x_i - \bar{x})\Omega^{-1}(x_i - \bar{x})' \tag{3}$$

The relevance of an observation x_i is equal to the sum of its multivariate similarity and its informativeness:

$$Relevance(x_i) = Similarity(x_i, x_t) + Informativeness(x_i) \tag{4}$$

To summarize, similarity equals the negative of the Mahalanobis distance of a prior observation of x_i from its current observation x_t. Informativeness equals the Mahalanobis distance of x_i from its historical average. Relevance equals the sum of similarity and informativeness. In other words, prior periods that are like the current period but are different from the historical average are more relevant than those that are not.

These notions are illustrated in Exhibit 1, which shows a scatter plot of two hypothetical variables. These variables both have a standard deviation of 5% and are 50% correlated. The gray dot represents current values, and the two black dots represent values in two prior periods. To determine each prior observation's relevance to the current observation, we first measure their multivariate similarity to the current observation. In this example, both prior observations have the same multivariate similarity as each other with respect to the current observation (0.5). Next, we quantify each prior observation's informativeness as its multivariate distance from the origin. All else equal, more extreme observations, such as point B, are more informative than moderate observations, such as point A. Finally, each prior observation's overall relevance is summarized as the sum of its similarity to the current observation and its informativeness. In this example, points A and B are equally similar to the current environment; however, B is more relevant because it is more informative.

LINEAR REGRESSION AS A RELEVANCE-WEIGHTED AVERAGE OF THE DEPENDENT VARIABLE

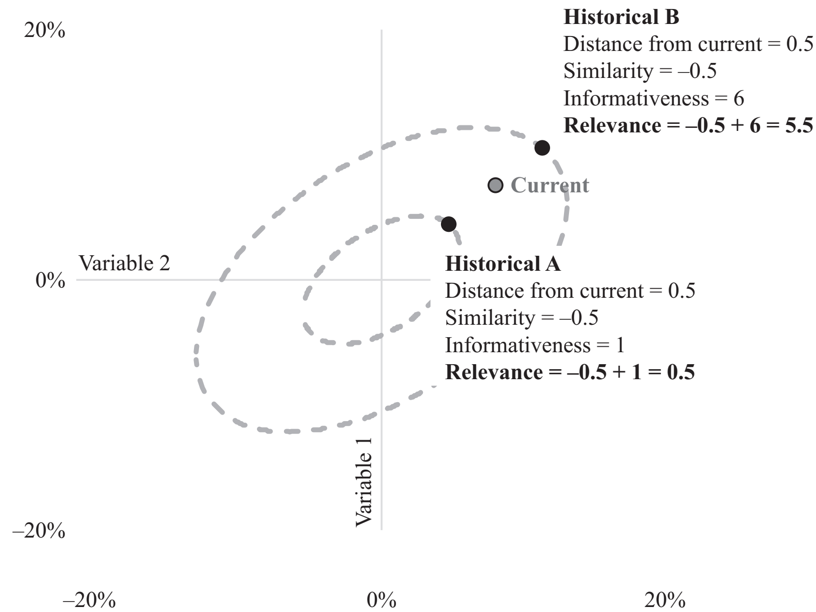
It is mathematically equivalent to interpret $\hat{y}_t$ from a fitted ordinary least squares regression equation as the weighted average of the prior y_i's, in which the weights equal the relevance of the prior x_i's.

[1] The Mahalanobis distance was introduced in Mahalanobis (1927) and Mahalanobis (1936).

[2] For further discussion of noise-driven versus event-driven observations and their relationship to estimating risk, see Chow et al. (1999).

E X H I B I T 1

Similarity, Informativeness, and Relevance



Let's begin by restating Equation 4 using the notation of Equations 2 and 3 and then simplifying, as shown in Equations 5 and 6. For notational simplicity we will assume that X is shifted to have a mean of zero.

$$Relevance(x_i) = -(x_i - x_t)\Omega^{-1}(x_i - x_t)' + x_i\Omega^{-1}x_i' \tag{5}$$

$$Relevance(x_i) = 2x_t\Omega^{-1}x_i' - x_t\Omega^{-1}x_t' \tag{6}$$

We assert that the prediction $\hat{y}_t$ from a fitted regression equation is equivalent to a relevance-weighted average of prior observations for y_i times a simple scalar multiple of 1/2:

$$\hat{y}_t = \frac{1}{2}\cdot\frac{1}{N}\sum_{i=1}^{N} Relevance(x_i)y_i \tag{7}$$

$$\hat{y}_t = \frac{1}{2N}\sum_{i=1}^{N} 2x_t\Omega^{-1}x_i'\,y_i - \frac{1}{2}x_t\Omega^{-1}x_t'\overline{y} \tag{8}$$

Next, let's assume, without loss of generality for our purposes, that the observed y_i's are shifted to have a mean value of zero, which causes the final term to drop out and yields

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

The expression for the covariance matrix (remembering that we have assumed means of zero for each component of X) is given by

$$\Omega = \frac{1}{N}X'X \tag{10}$$

The inverse of the covariance matrix is given by

$$\Omega^{-1} = N(X'X)^{-1} \tag{11}$$

Substituting this value into the expression for $\hat{y}_t$, we have

$$\hat{y}_t = \sum_{i=1}^{N} x_t(X'X)^{-1}x_i'\,y_i \tag{12}$$

By expressing Equation 12 in standard matrix notation, we have

$$\hat{y}_t = x_t(X'X)^{-1}X'Y \tag{13}$$

EXHIBIT 2
Equivalence of Fitted Regression and Relevance-Weighted Dependent Variables

Panel A: Inputs

	(A)	(B)	(C)	(D)	(E)	(F)
Observation (i)	**Variable 1**	**Variable 2**	**Asset Return**	**De-meaned Variable 1 (X_1)**	**De-meaned Variable 2 (X_2)**	**De-meaned Asset Return (Y)**
1	–0.7%	–1.3%	20.6%	–0.8%	–1.1%	20.6%
2	1.5%	–0.2%	–2.2%	1.4%	–0.1%	–2.2%
3	–0.1%	–0.9%	–9.9%	–0.2%	–0.7%	–9.9%
4	–0.5%	0.0%	–0.2%	–0.7%	0.2%	–0.2%
5	–0.1%	1.4%	7.4%	–0.2%	1.6%	7.4%
6	0.7%	0.0%	–15.8%	0.6%	0.2%	–15.8%

Panel B: Standard Approach (based on de-meaned variables)

	(G)	(H)	
	X_1	X_2	**Prediction ($\hat{y}$) for $t = 6$**
Beta	–8.0	–0.6	
Values for $t = 6$	0.6%	0.2%	**Sum of weighted Xs**
Beta-weighted X	–4.8%	–0.1%	**–4.9%**

Panel C: Relevance-Weighted Approach (based on de-meaned variables)

	(I)	(J)	(K)	(L)	(M)	(N)
Observation (i)	**Distance from $t = 6$ $(x_i - x_t)\Omega^{-1}(x_i - x_t)'$**	**Similarity –1 * (I)**	**Informativeness $x_i\Omega^{-1}x_i'$**	**Relevance (J) + (K)**	**Scaled Relevance (L)/[2*(n_obs–1)]**	**Relevance-Weighted Y (M) * (F)**
1	4.2	–4.2	2.0	–2.1	–0.21	–4.4%
2	1.1	–1.1	2.9	1.8	0.18	–0.4%
3	1.6	–1.6	0.6	–1.0	–0.10	1.0%
4	2.4	–2.4	0.8	–1.6	–0.16	0.0%
5	3.6	–3.6	3.2	–0.4	–0.04	–0.3%
6	0.0	0.0	0.5	0.5	0.05	–0.8%
						Sum of weighted Ys
					Prediction ($\hat{y}$) for $t = 6$	**–4.9%**

This leads us to the familiar standard solution for generating a prediction from a fitted linear regression:

$$\hat{y}_t = x_t\beta' \tag{14}$$

A more general case permits Y to have a non-zero mean, $\overline{y} = \frac{1}{N}\Sigma_{i=1}^{N} y_i$. In this case, the formula becomes

$$\hat{y}_t = \overline{y} + \frac{1}{2N}\sum_{i=1}^{N} Relevance(x_i)(y_i - \overline{y}) \tag{15}$$

We next demonstrate this equivalence with a simple example. We first regress values for a hypothetical dependent variable on two hypothetical independent variables. The inputs are shown in the top panel of Exhibit 2. For convenience, we de-mean the variables before fitting the regression. The middle panel applies the beta coefficients to the current-period observations ($t = 6$) to yield the prediction for the dependent variable $\hat{y}$ of −4.9%. The final panel illustrates the equivalent approach of weighting the y_i's by their relevance to arrive at $\hat{y}$. The process is as follows:

1. First, we measure the multivariate distance of each observation from the current observation (time $t = 6$) based on their respective x_i's (column I). The opposite (negative) of these distances represents the similarity of each historical observation to the current observation (column J).
2. We then quantify the informativeness of each historical observation as its multivariate distance from the origin (column K). All else equal, extreme

observations (those that lie far from the origin) are more informative than moderate observations (those that lie close to the origin).

3. We next sum each observation's similarity to the current observation and its informativeness to derive an overall relevance score (column L).
4. Finally, to generate a prediction for time $t = 6$, we scale the relevance scores (column M) and multiply them by their respective Y values (column N), the sum of which is the predicted value $\hat{y}$.

Note that this procedure generates a prediction, −4.9%, that is identical to the regression approach illustrated in the middle panel.

CREATING A SUBSAMPLE OF RELEVANT OBSERVATIONS

The equivalence of a fitted regression prediction and a relevance-weighted average of the past values of the dependent variable offers an intriguing insight into the workings of linear regression. It reveals the implicit assumption of linear regression, which is that whatever occurred during similar periods in history will recur, and whatever occurred during dissimilar periods of history will occur but in the opposite direction. This insight invites the question of whether data from similar and dissimilar past observations are equally useful. In other words, should we be more inclined to extrapolate observations from similar past periods than from dissimilar past periods (but in the opposite direction)?

We can address this question by partitioning our sample into subsamples that pass relevance thresholds, considering both similarity and informativeness. We then apply our relevance-weighted methodology to these subsamples to predict the dependent variable:

$$\hat{y}_t = \overline{y} + \frac{1}{2n}\sum\nolimits_{relevant\ i's}^{n} Relevance(x_i)(y_i - \overline{y}) \tag{16}$$

In Equation 16, the summation only occurs over the subsample of historical observations that meet or exceed a chosen relevance threshold, which we denote as *relevant i's*. Likewise, the mean is evaluated on this same subsample, $\overline{y} = \frac{1}{n}\sum\nolimits_{relevant\ i's}^{n} y_i$. In contrast to Equation 15, which averages across all N observations in the full sample, Equation 16 averages across the total number of observations in the subset, n. This formula converges to two intuitive edge cases. If we include all N observations, the estimate is equal to that of a standard linear regression. If we include only a single data point x_i that is precisely equal to x_t, the prediction equals y_i for that observation.

It is important to note that partial sample regression is not equivalent to running a new regression on the identified subsamples because linear regression infers relevance from the full sample. If we were to run a new regression on a subsample of relevant observations, that regression would only "know" about these relevant observations. This new regression would fail to recognize that all the observations are relevant to the conditions we are predicting and therefore would produce a biased result. In other words, we must censor the predictive contributions of irrelevant observations after we assess their relevance and without changing our assessment of their relevance.

It is easy to show using simulation that if the data come from a single stationary multivariate normal distribution, the expected predictions from each subsample will be equivalent. The question we wish to address is whether this prediction holds empirically. Stated differently, can we improve the quality of our prediction by using more relevant prior observations? Before we proceed to our empirical analysis, it might be helpful to place our innovation in the context of previous research.

Relationship to Previous Research

Our partial sample regression methodology may be viewed as an application of the so-called Nadaraya–Watson kernel regression. Nadaraya (1964) and Watson (1964) independently proposed an estimator for predicting $\hat{y}$ that consists of a kernel-weighted average of the historical data points Y. The kernel $K_h(x_t, x_i)$ refers to a type of function that assigns values based on the distance between two points. Points that are close together receive greater value, whereas those that are more distant are assigned lower values. The assignment of these values, and thus the way they decay with increasing distance between points, is determined by a function or parameter h called the *bandwidth*. Kernel regression is often called *kernel smoothing* because it is a refinement of k-nearest-neighbors estimation, in which $\hat{y}$ is taken to be the equally weighted average of Y for a subset of the closest data points in X.

A wide variety of kernel functions have been proposed in the literature. They vary in how distance is measured between two points and in the shape of the weight decay applied to larger distances. It is most common to use the simple Euclidean distance to assess the similarity between observations and to use various functional forms of a bell-shaped curve to de-emphasize larger distances.

To the best of our knowledge, researchers and practitioners do not commonly use the Mahalanobis distance in this context. Moreover, we are not aware of any previous research documenting the fact that a Mahalanobis-based kernel exactly reproduces the predictions from ordinary least squares linear regression. We believe that this connection is powerful, not only to justify extrapolation of historically relevant observations but also to lend further intuition to the results of linear regression analysis and its extension to our relevance-weighted average methodology.

We should also note that our approach is different from weighted least squares regression, which uses fixed weights regardless of the data point being predicted and applies the weights to calculate the covariance matrix among predictors; in contrast, we use the full historical covariance matrix to measure distance. Moreover, our approach is different from performing separate regressions on subsamples of the most relevant observations; in a separate regression approach, the covariance matrix used for estimation would also be based on the subsample, whereas we use the full-sample covariance matrix. Lastly, other approaches to kernel regression only consider similarity, whereas our approach considers both similarity and informativeness.

EMPIRICAL ANALYSIS

To illustrate our methodology, we use economic variables to predict factor performance. Considerable research suggests that factor performance can be explained by economic variables. Recent articles on this topic include those by Bender et al. (2018), Amenc et al. (2019), and Fergis et al. (2019). Each of these articles includes a summary of previous research. However, there is widespread disagreement on which economic variables are important and in what way they affect factor performance. Many researchers have noted that the effects are not necessarily constant over time: Some relationships seem to be stronger during particular calendar periods or during particular economic and financial regimes. We wish to determine whether our methodology captures the time-varying nature of these relationships and therefore offers better-informed predictions of factor returns. With our approach, we do not need to distinguish between a predicting variable and a conditioning variable. Relevance and prediction are evaluated simultaneously across the variables we choose for X.

Data and Calibration

We consider seven long–short US equity factors[3]:

- Equity (S&P 500 minus Bloomberg US Treasuries)
- Size (Kenneth French's small minus big factor)
- Value (Kenneth French's high minus low factor)
- Profitability (Kenneth French's robust minus weak factor)
- Investment (Kenneth French's conservative minus weak factor)
- Momentum (Kenneth French's momentum factor)
- Volatility (Kenneth French's low-volatility [bottom quintile] portfolio minus high-volatility [top quintile] portfolio)[4]

We use eight US economic variables as our predictors. Specifically, we include the level (defined in parentheses) and one-year change of the following variables[5]:

- Economic growth (year-over-year percentage change in real gross national product, seasonally adjusted)
- Unemployment rate (civilian unemployment rate, seasonally adjusted)
- Inflation (year-over-year percentage change in the Consumer Price Index, seasonally adjusted)
- Credit spreads (Moody's BAA rate—10-year Treasury constant maturity rate)

Results

For each factor, we regress historical returns on lagged observations of the economic variables.[6] The dependent variable in each regression is the past year's

[3] We obtain the S&P 500 Index and Bloomberg US Treasuries from Datastream. All Kenneth French data are obtained from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

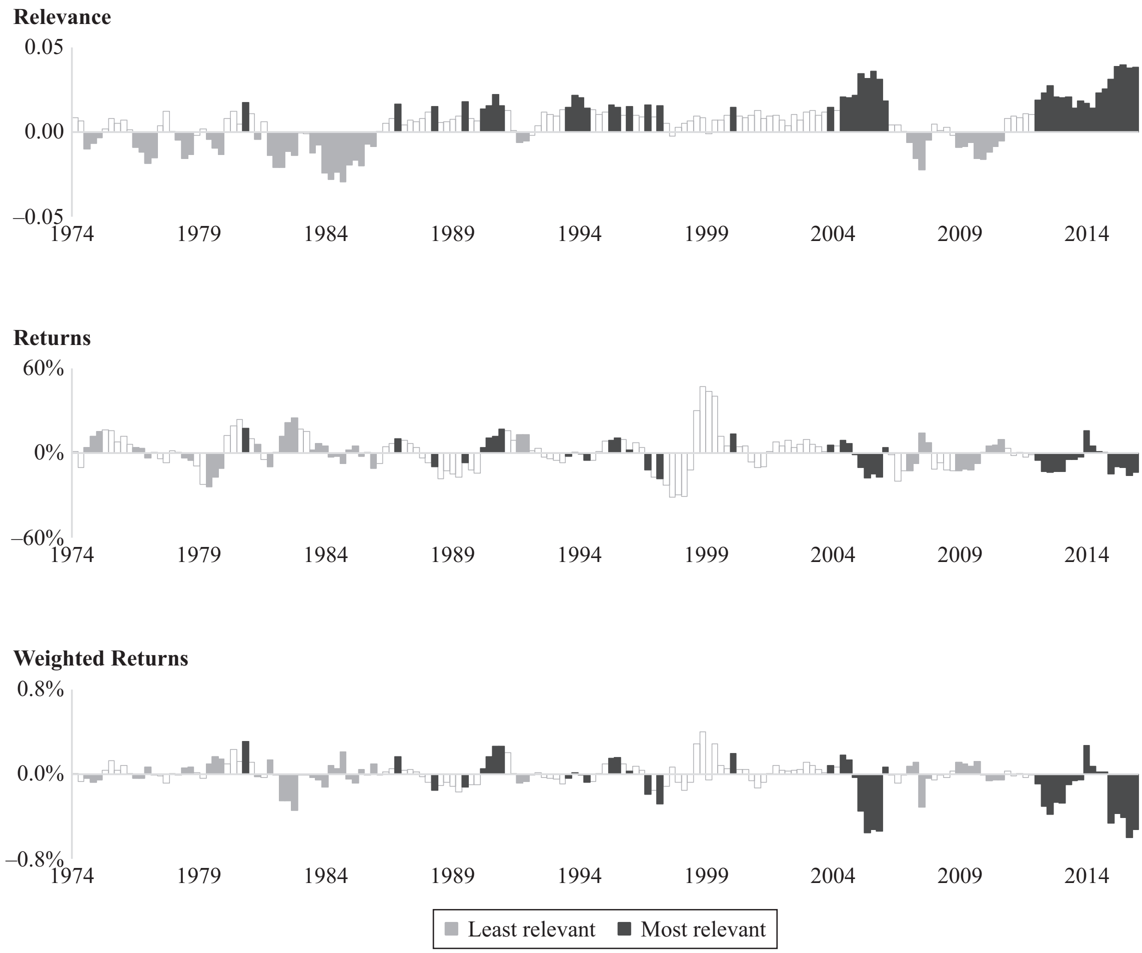
[4] We lever the low-volatility portfolio such that it has the same volatility as the high-volatility portfolio (based on trailing 60-month returns). We assume borrowing at the risk-free rate.

[5] All economic data are sourced from the Federal Reserve of St. Louis Economic Data (FRED) online library.

[6] We use log returns for Y and de-mean the X variables before running the regression.

EXHIBIT 3

Relevance-Weighted Value Returns, June 1974–December 2018



return as of each quarter end. The independent variables are one-year-lagged levels and changes in the economic variables from the year preceding the returns. For each period's observation (x_t), we generate relevance-weighted predictions $\hat{y}$ using the methodology described previously. We do this for two subsamples: the 25% of observations least relevant to x_t and the 25% most relevant to x_t. Finally, we evaluate the quality of our predictions for three approaches: a fitted regression equation based on the full sample, a relevance-weighted average applied to the 25% least relevant periods, and a relevance-weighted average applied to the 25% most relevant periods. As our figure of merit, we use the correlation of the predicted returns with the actual returns. We use simulation to determine the statistical significance of the difference between the partial sample (most relevant) and full sample regression models.[7]

[7] To generate a P-value of asymmetry, we compare the difference in correlation of the partial sample regression and full sample regression to a simulated distribution of differences that assumes all data points equally reflect the true relationship between Y and X. Specifically, for each factor, we simulate observations for the hypothetical X variables by (1) drawing quarterly values from the multivariate normal distribution that characterizes the variables at a quarterly frequency and (2) summing the simulated quarterly variables over rolling four-quarter periods. This mimics the overlapping nature of our observations. To generate the simulated Y variable observations, we apply the beta coefficients from the actual full sample regression to the simulated X variables and add simulated error terms drawn from the distribution of actual errors. We repeat this process 1,000 times to generate a distribution of correlation differences between the partial sample and full sample regression

EXHIBIT 4
Correlations of Model Returns and Actual Returns

	Full Regression	Filtered: Least Relevant	Filtered: Most Relevant	Difference: Most Relevant-Full	P-value of Difference
Equity	33.5%	–14.9%	62.0%	28.5%	0.04
Size	55.5%	–24.1%	63.8%	8.4%	0.58
Value	36.9%	–5.5%	61.4%	24.4%	0.10
Profitability	22.2%	–4.4%	53.5%	31.4%	0.09
Investment	41.5%	–10.4%	65.2%	23.7%	0.05
Momentum	62.2%	–21.8%	70.0%	7.8%	0.35
Volatility	35.7%	–6.5%	57.1%	21.4%	0.21

Note: P-values reflect the null hypothesis that all data points equally reflect the true relationship between Y and X.

Before we proceed to the results, we present a visualization of how we weight the value factor returns, as an example, based on the relevance of historical periods to produce predictions of the return to the value factor in the final period.[8]

The black bars in the top panel of Exhibit 3 represent the 25% of periods that are most relevant based on the economic variables described earlier. The gray bars represent the 25% that are least relevant. The white bars represent the remaining periods. The middle panel shows the returns of the value factor as defined earlier, and the bottom panel, which is the product of the top two panels, gives the relevance-weighted returns.

This depiction allows us to observe which periods make more important contributions in predicting the dependent variable. The full-sample prediction, which comes from the fitted regression equation, is −3.4%. The least relevant observations predict a return of +1.1% based on the relevance-weighted average approach. The most relevant observations, also based on the relevance-weighted average approach, predict that the value factor will return −10.5% in the next period. It is clear that these different samples produce different outcomes. We next seek to determine which approach produces the most reliable outcome.

Exhibit 4 shows the results based on data from June 1974 through December 2018. For each factor, the difference in fit between the partial sample and full sample regression models is positive and in most cases statistically significant. Moreover, return predictions derived from the least relevant observations are negatively correlated with actual returns. These results strongly suggest that relevant economic environments contain more information about future factor performance than less relevant environments.

models. The P-value equals the fraction of simulated differences that exceeds the actual difference.

[8] For purposes of illustration, we shift X and Y to have mean values of zero. This allows us to exclude the second term in Equation 6 because it is merely a constant that multiplies the average of Y.

SUMMARY

Financial analysts employ regression analysis in a variety of settings to predict variables of interest. Those trained in classical statistics commonly assume that the quality of a prediction improves with sample size because larger samples tend to produce less noisy results than smaller samples. But this is not always the case. We argue that, in some cases, one can produce more reliable predictions by using subsamples of the original data in which less relevant observations have been censored. We define relevance in a mathematically precise way as the sum of multivariate similarity and informativeness. Given this precise definition of relevance, we show that the prediction from a fitted regression equation is equivalent to the weighted average of the past values of the dependent variable, in which the weights are the relevance of the independent variables. We show how to use this equivalence to derive predictions from a subsample of relevant observations and apply this methodology to predict factor returns from economic variables. Our results suggest that in certain settings we produce more reliable predictions by taking a weighted average of past values for the dependent variable based on a sample that has been censored for relevance than by applying regression analysis to the full sample. We do not argue that our results apply generally. However, for settings in which the data are unlikely to be stationary and multivariate normal and in which there may be significant disparity in the relevance of historical observations, we strongly recommend our methodology.

ACKNOWLEDGMENTS

We thank Robin Greenwood, Joseph Howlett and Bruce Turkington for helpful comments.

REFERENCES

Amenc, N., M. Esakia, F. Goltz, and B. Luyten. 2019. "Macroeconomic Risks in Equity Factor Investing." *The Journal of Portfolio Management* 45 (6): 39–60.

Bender, J., X. Sun, R. Thomas, and V. Zdorovtsov. 2018. "Promises and Pitfalls of Factor Timing." *The Journal of Portfolio Management* 44 (4): 79–92.

Chow, G., E. Jacquier, M. Kritzman, and K. Lowry. 1999. "Optimal Portfolios in Good Times and Bad." *Financial Analysts Journal* 55 (3): 65–71.

Fergis, K., K. Gallagher, P. Hodges, and K. Hogan. 2019. "Defensive Factor Timing." *The Journal of Portfolio Management* 45 (3): 50–68.

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

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

Nadaraya, E. A. 1964. "On Estimating Regression." *Theory of Probability and Its Applications* 9 (1): 141–142.

Watson, G. S. 1964. "Smooth Regression Analysis." *Sankhya: The Indian Journal of Statistics Series A* 26 (4): 359–372.

ADDITIONAL READING

Macroeconomic Risks in Equity Factor Investing

Noël Amenc, Mikheil Esakia, Felix Goltz, and Ben Luyten

The Journal of Portfolio Management

https://jpm.pm-research.com/content/45/6/39

ABSTRACT:*There is a consensus that equity factors are cyclical and depend on macroeconomic conditions. To build well-diversified portfolios of factors, one needs to account for the fact that different factors may have similar dependencies on macroeconomic conditions. The authors provide a protocol for selecting relevant macroeconomic state variables that reflect changes in expectations about the aggregate economy. They show that returns of standard equity factors depend significantly on such state variables. Factor returns also depend on aggregate macroeconomic regimes reflecting good and bad times. These macroeconomic risks have strong portfolio implications. For example, some equity factors depend on interest rate risk. Investors who already have exposure to this risk through bond investments may increase loss risk when tilting to the wrong equity factors. The authors also show that standard multifactor allocations do not sufficiently address macroeconomic conditionality. Combining factors may not reduce macroeconomic risks even for factors with low correlation. Understanding macroeconomic risks is a prerequisite both for risk transparency and for improving diversification of equity factor investments.*

The Promises and Pitfalls of Factor Timing

Jennifer Bender, Xiaole Sun, Ric Thomas, and Volodymyr Zdorovtsov

The Journal of Portfolio Management

https://jpm.pm-research.com/content/44/4/79

ABSTRACT: *The potential to dynamically allocate across factors, or factor timing, has been an area of academic and practitioner research for decades. In this article, the authors revisit the promises of factor timing, documenting the historical linkages between equity factor performance and different groupings of predictors: sentiment, valuation, trend, economic conditions, and financial conditions. The authors highlight that different predictors are more relevant for certain horizons, so the horizon is critical in factor timing. They also argue there are significant pitfalls with factor timing as well. The difficulty of timing factors has been well documented, given the uncertainty of exogenous elements affecting their behavior and the complexity of the underlying relationships. Most importantly, the underlying causal links are time varying. In addition, these relationships are observed with the benefit of hindsight and thus suffer from the age-old problem of data mining. Despite these caveats, the authors believe that at the margin it is possible to time certain elements that can add value and improve outcomes.*

Enhanced Scenario Analysis

Megan Czasonis, Mark Kritzman,
Baykan Pamir, and David Turkington
The Journal of Portfolio Management
https://jpm.pm-research.com/content/46/4/69

ABSTRACT: *Investors have long relied on scenario analysis as an alternative to mean–variance analysis to help them construct portfolios. Even though mean–variance analysis accounts for all potential scenarios, many investors find it difficult to implement because it requires them to specify statistical features of asset classes that are often unintuitive and difficult to estimate. Scenario analysis, by contrast, requires only that investors specify a small set of potential outcomes as projections of economic variables and assign probabilities to their occurrence. It is, therefore, more intuitive than mean–variance analysis, but it is highly subjective. In this article, the authors propose to replace the subjective elements of scenario analysis with a robust statistical process. They use a multivariate measure of statistical distance to estimate probabilities of prospective scenarios. Next, they construct portfolios that maximize utility for investors with different risk preferences. Last, the authors introduce a procedure for minimally modifying scenarios to render them consistent with prespecified views about their probabilities of occurrence.*