# PORTFOLIO CONSTRUCTION WHEN REGIMES ARE AMBIGUOUS

THIS VERSION

APRIL 13, 2023

**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.
100 Main Street, Cambridge MA, 02142
kritzman@mit.edu
617-901-4455

**Cel Kulasekaran** is managing director of research at Windham Capital Management in Boston, MA.
1 Federal Street, Boston, MA, 02110
cel@windhamcapital.com
617-419-3921

**David Turkington** is senior managing director and head of State Street Associates in Cambridge, MA.
140 Mt Auburn Street, Cambridge MA, 02138
dturkington@statestreet.com
617-955-7679

Abstract

Investors sometimes have strong convictions that a distinctive economic regime will prevail in the period ahead and therefore would like to form a portfolio that reflects the expected returns, standard deviations, and correlations of assets during such a regime. To do so, they typically isolate a subsample of returns in which a regime indicator, such as the rate of economic growth, is above or below a chosen threshold and estimate expected returns, standard deviations, and correlations by equally weighting the observations within the subsample. This approach assumes that every observation within the regime subsample is equally important to forming the estimates whether an observation coincides with a growth rate that is far from the threshold or one that is only marginally distant from the threshold. Moreover, with this approach it is problematic to describe a regime by more than a single indicator because there is no non-arbitrary way to combine the indicators and because the addition of indicators increases the likelihood of producing an empty or overly sparse subsample. The authors apply a new concept called relevance to estimate regime-specific expected returns, standard deviations, and correlations. Their relevance-based approach explicitly accounts for the importance of an observation to forming an estimate, and it seamlessly enables the inclusion of multiple regime indicators in a principled way.

# PORTFOLIO CONSTRUCTION WHEN REGIMES ARE AMBIGUOUS

Investors sometimes prefer to form portfolios that they expect will perform favorably during a particular economic regime such as a challenging economic growth regime. To do so, they isolate a subsample of returns that prevailed during periods when the regime indicator fell below a specified threshold and estimate expected returns, standard deviations, and correlations by equally weighting the observations within the regime subsample. They therefore assume that an observation for which economic growth was only marginally below the regime threshold is equally as important to forming an estimate as an observation for which growth was far below the regime threshold. In addition to this simplistic binary view of an observation's importance, there is no non-arbitrary way to combine multiple regime indicators. Moreover, the inclusion of additional indicators increases the likelihood of producing an empty or overly sparse subsample. We propose an alternative approach for forming regime sensitive portfolios based on the concept of relevance, which gives a mathematically precise and theoretically justified measure of the importance of an observation to forming an estimate. Our relevance-based approach for estimating regime specific expected returns, standard deviations, and correlations also seamlessly accommodates multiple regime indicators in a principled way.

We begin by introducing the concept of regimes that are identified statistically. Next, we define relevance and describe how it is used to form estimates of regime specific return and risk. We then provide an illustration of our relevance-based approach for forming estimates and compare it to the conventional approach, given a regime that is described by a single indicator. Next, we show how to extend our relevance-based approach for forming a regime sensitive portfolio given a regime that is defined by more than a single indicator. We conclude with a summary.

**Non-binary regimes**

Investors often define a regime as a collection of periods for which an indicator variable falls above or below a fixed threshold. This definition is binary because each period is either included or not included in the regime. It may be more useful, however, to define regimes by specifying the prototypical circumstances that characterize the regime and allowing statistical techniques to determine the relevance of each period to those circumstances. This definition of regimes is non-binary because each period may contribute a different degree of information about the regime. In other words, regime identification is ambiguous. In this more flexible approach, we use any periods that are relevant to the prototypical regime circumstances to inform our estimates of return and risk. Moreover, the same period may be used to inform the estimates of multiple regimes. The process of specifying prototypical circumstances to characterize regimes is intuitive, and it is related to the common practice of scenario analysis. The circumstance that defines a regime is a single value in the case of one indicator, or a vector of values in the case of multiple indicators. Next, we explain how to measure the relevance of each period to a chosen circumstance.

**Relevance**

Relevance measures the importance of an observation to forming an estimate. It is composed of two components, similarity and informativeness, as shown by Equation 1. If a regime is defined by a single indicator, similarity and informativeness are measured as squared z-scores as shown by Equations 2, 3, and 4.

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

$$sim(x_i, x_t) = -\frac{1}{2}(x_i - x_t)^2/\sigma_x^2 \tag{2}$$

$$info(x_i, \bar{x}) = (x_i - \bar{x})^2/\sigma_x^2 \tag{3}$$

$$info(x_t, \bar{x}) = (x_t - \bar{x})^2/\sigma_x^2 \tag{4}$$

In these equations, $x_i$ is the value of the regime indicator for a prior observation, $x_t$ is the value of the regime indicator that characterizes the prospective regime, $\bar{x}$ is the average of all the observations including the current values, and $\sigma_x$ is the standard deviation of all the $x_i$s.

If we instead define a regime by more than a single indicator, we must use the Mahalanobis distance to measure similarity and informativeness as shown in Equations 5, 6, and 7.

$$sim(x_i, x_t) = -\frac{1}{2}(x_i - x_t)\Omega^{-1}(x_i - x_t)' \tag{5}$$

$$info(x_i, \bar{x}) = (x_i - \bar{x})\Omega^{-1}(x_i - \bar{x})' \tag{6}$$

$$info(x_t, \bar{x}) = (x_t - \bar{x})\Omega^{-1}(x_t - \bar{x})' \tag{7}$$

In these equations, $x_i$ is a vector of the values of the regime indicators for a prior observation, $x_t$ is a vector of the values of the regime indicators that are expected to prevail during the prospective regime, $\bar{x}$ is the average of all the observations including the current values, and $\Omega^{-1}$ is the inverse covariance matrix of all the $x_i$s. The vector $(x_i - x_t)$ measures how distant the observations are independently from their expected values during the regime. By multiplying this vector by the inverse of the covariance matrix, we capture the interaction of the observations, and at the same time we standardize the distances by dividing by variance. By multiplying this product by the transpose of the vector $(x_i - x_t)$ we collapse the outcome into a single number.

Notice that for our measure of similarity, whether we use a squared z-score or a Mahalanobis distance, we multiply by negative ½. The negative sign converts a measure of distance into a measure of similarity. We multiply by ½ because the distances between two observations (a prior observation and the regime value observation) have the potential to be twice as large as the observations' distances from the average of all observations. When we measure informativeness, we retain its positive sign, and we have no need to multiply by ½. By measuring informativeness as a difference from average, we are claiming that unusual observations contain more information than common observations, which follows from Claude Shannon's information theory.[1] Finally, note that whether we measure relevance from squared z-scores or Mahalanobis distances, we also measure the unusualness of the current observation. We do so to center our measure of relevance on zero. All else being equal, observations that are similar to the circumstances that characterize a regime but different from average circumstances are more relevant than those that are not.

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

We can also justify the non-arbitrariness of relevance in the following sense. A relevance weighted average of prior outcomes yields an estimate that is precisely equivalent to the estimate that results from linear regression analysis when applied across the full sample of observations. We provide details of this equivalence in the Appendix. Even though we intend to use only a subset of the most relevant observations to construct our estimates, this equivalence provides an important theoretical foundation for our measurement of statistical relevance.

We form our estimates of expected returns, standard deviations, and correlations as relevance weighted averages of the past values of these outcomes for the observations within a regime subsample that is determined by relevance. This approach enables us to capture the realistic ambiguity of an observation's association with a regime, rather than naively asserting that an observation is unambiguously within a regime or unambiguously outside a regime.

Our relevance-based approach to forming estimates is a theoretically grounded refinement to kernel regression. Kernel regression forms an estimate as a weighted average of local observations, by applying a Gaussian decay to normalized Euclidean distances (Gaussian kernel) to compute the weight of each observation. Our relevance-based approach, by contrast, uses the Mahalanobis distance instead of the Euclidean distance to measure nearness, and it adds the element of informativeness along with nearness to determine relevance. Forming an estimate from a subsample of the most relevant observations is called partial sample regression, which is described by Equation 8.

$$\hat{y}_t = \bar{y} + \frac{\lambda^2}{n-1} \sum_{i \in R_{sub}} r_{it}(y_i - \bar{y}) \tag{8}$$

In Equation 8, $i \in R_{sub}$ indicates the set of observations $i$ that are contained within the regime subsample, $R_{sub}$. The term $\frac{\lambda^2}{n-1}$ compensates for a bias that would otherwise arise from focusing on a small subsample of observations. Equation 8 tilts equal weights toward observations that are more relevant.

Equivalently, we can express the estimates from Equation 8 as a weighted average of $y_i$ outcomes in which the weights sum to 1.

$$\hat{y}_t = \sum_{i=1}^{N} w_{it} y_i \tag{9}$$

$$w_{it} = \frac{1}{N} + \frac{\lambda^2}{n-1}(\delta(r_{it})r_{it} - \varphi \bar{r}_{sub}) \tag{10}$$

In Equation 10, $\delta(r_{it})$ is a censoring function that equals 1 if $r_{it} \geq r^*$ and 0 otherwise. The threshold $r^*$ determines what fraction of observations to consider for a given regime's estimates. It is useful to impose such a threshold to the extent that the least relevant observations are less reliable than the most relevant observations, which is often the case.[2] For notational concision we write the number of observations for which $\delta(r_{it}) = 1$ as $n = \sum_i \delta(r_{it})$ and the proportion of all observations for which $\delta(r_{it}) = 1$ as $\varphi = \frac{n}{N}$. In addition, we write the subsample average of relevance over the retained observations as $\bar{r}_{sub} = \frac{1}{n}\sum_i \delta(r_{it})r_{it}$. The adjustment factor is defined as $\lambda^2 = \sigma^2_{r,full}/\sigma^2_{r,partial} = \frac{1}{N-1}\sum_i r^2_{it} \,/\, \frac{1}{n-1}\sum_i \delta(r_{it})r^2_{it}$.

In addition to its theoretical justification and its accommodation of multiple regime indicators, our relevance-based approach for defining a period's regime exposure has an important practical advantage: it enables us to consider regimes that have not yet occurred historically yet are plausible looking forward.

Equation 8 is completely general in the sense that it provides an estimate for any outcome, $y$. In the context of regime specific portfolio construction, the first quantity we must estimate is the vector of conditional expected returns for the assets in our investment universe. For this task we define $y$ as the return of a chosen asset in each period which we repeat for every asset. Note that in the absence of predictive $X$ variables, the estimate for each asset will equal its full sample average return.

The second quantity we must estimate is the covariance matrix of asset returns. Let us start by considering the diagonal elements of the covariance matrix, which represent the variance of returns for each asset. To build intuition for this process, it is helpful to view the variance of an asset as an estimate of the expected squared deviation of that asset's return from its average. In the absence of predictive $X$ variables, the estimate of variance will equal the simple average of squared deviations from the full sample expected return. In the presence of predictive $X$ variables, the estimate will equal a weighted average of squared deviations from the (weighted) conditional expected return. Therefore, to estimate the variance of an asset we define $y$ as the squared deviation from the estimated expected return of the subsample.

We estimate the off-diagonal elements of the covariance matrix in a similar fashion, setting $y$ equal to one asset's return deviation from its conditional average multiplied by a second asset's return deviation from its conditional average. We use Equation 8 to estimate the conditional pairwise covariance for every pair of assets. The resulting covariance matrix defines both the volatilities and correlations of the assets specific to the chosen regime.

We next illustrate how our relevance-based approach compares to the conventional approach for estimating regime specific expected returns, standard deviations, and correlations, given a regime that is defined by a single indicator. As we mentioned earlier, when we define a regime based on only a single indicator, we measure relevance by square z-scores rather than Mahalanobis distances.

**Relevance-based prediction versus conventional approach given a single regime indicator**

We illustrate our relevance-based approach to forming regime specific estimates of expected returns, standard deviations, and correlations for a low economic growth regime in which annual real GDP growth is less than -2.0%.

We consider six asset classes for our analysis, and we observe their returns yearly from 1974 through 2022.

| | |
|---|---|
| US Equities | S&P 500 |
| Foreign Equities | MSCI World ex USA |
| US Treasuries | Bloomberg Barclays Treasury Bond Index |
| Corporate Bonds | Bloomberg Barclays Corporate Bond Index |
| Commodities | Bloomberg Commodity Index |
| Risk Free Asset | Risk-free Rate from Kenneth French's data website |

We form our relevance-based predictions as follows:

1. We specify a regime indicator (annual real GDP growth) and a threshold (-2.0%) to isolate the observations we use to form the predictions.
2. We calculate the squared z-scores of the observations relative to their historical average.
3. We calculate the squared z-scores of the observations relative to -2% real GDP growth.
4. Using (2) and (3), we calculate the relevance of each observation.
5. We choose the 20% most relevant observations.
6. We estimate the expected returns and covariances using Equation 8.

The conventional approach of equally weighting observations requires only two steps for forming the predictions.

1. We specify a regime indicator (annual real GDP growth) and a threshold (-2.0%) to isolate the observations we use to form the predictions.
2. We estimate the expected returns and covariances by equally weighting the observations within the regime subsample.

Exhibit 1 shows the weights for the observations used to estimate the regime specific expected returns, standard deviations, and correlations.

Exhibit 1: Weights Used to Form Predictions

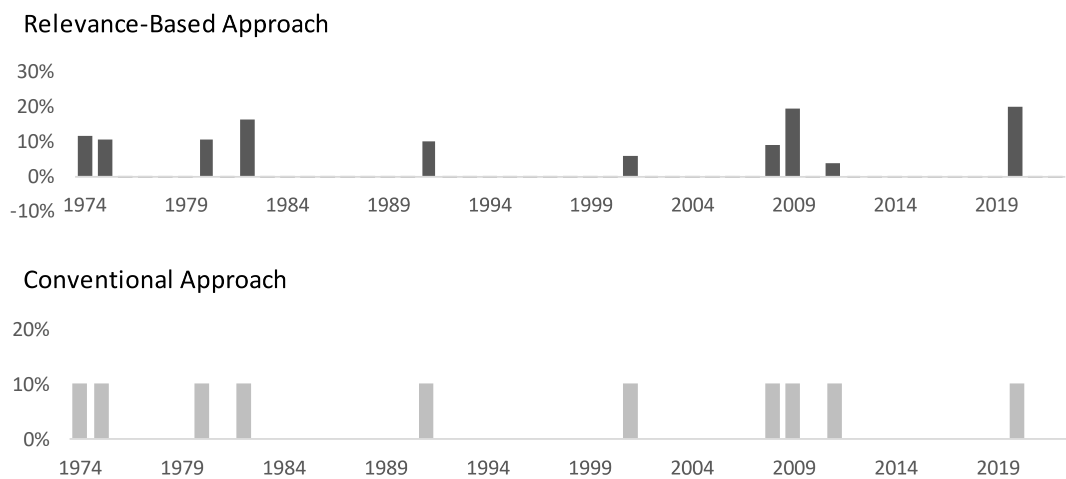


Exhibit 1 reveals that, coincidently, both the relevance-based approach and the conventional approach identify the same periods as constituting the low growth regime. However, the relevance-based approach for weighting the low-growth periods yields weights that vary quite starkly from the conventional approach's equal weights. All but one of the periods have weights that are below or above 10%, and most times they are significantly different.

Exhibit 2 shows the average returns, standard deviations, and correlations of the asset classes. The values in the top panel are computed by weighting the observations in the low growth regime by their relevance. The values in the middle panel are formed by weighting each observation in the low growth regime equally. The bottom panel shows these values for the full sample of observations assuming they are equally weighted. The relevance weighted approach gives values that are reasonably similar to the conventional equal weighting approach, though both approaches yield significant differences from the values obtained from the full sample of observations.

Exhibit 2: Regime Conditioned and Full Sample

Average Returns, Standard Deviations, and Correlations

| **Relevance Approach** | Average | Standard Deviation | Correlation | | | | | |
|---|---|---|---|---|---|---|---|---|
| US Equities | 13.9% | 25.2% | 1.00 | | | | | |
| Foreign Equities | 5.2% | 24.1% | 0.87 | 1.00 | | | | |
| Treasury Bonds | 10.3% | 9.6% | 0.01 | -0.46 | 1.00 | | | |
| Corporate Bonds | 13.9% | 13.8% | 0.69 | 0.35 | 0.53 | 1.00 | | |
| Commodities | -5.0% | 23.0% | 0.15 | 0.26 | -0.04 | 0.18 | 1.00 | |
| Cash | 4.8% | 4.4% | 0.13 | -0.08 | 0.59 | 0.23 | 0.54 | 1.00 |
| **Conventional Approach** | Average | Standard Deviation | Correlation | | | | | |
| US Equities | 9.3% | 25.2% | 1.00 | | | | | |
| Foreign Equities | 1.6% | 24.5% | 0.90 | 1.00 | | | | |
| Treasury Bonds | 9.9% | 7.7% | 0.05 | -0.33 | 1.00 | | | |
| Corporate Bonds | 10.9% | 12.4% | 0.67 | 0.38 | 0.54 | 1.00 | | |
| Commodities | -5.1% | 24.1% | 0.19 | 0.32 | -0.09 | 0.06 | 1.00 | |
| Cash | 4.7% | 4.0% | 0.23 | 0.16 | 0.43 | 0.18 | 0.48 | 1.00 |
| **Full Sample Estimates** | Average | Standard Deviation | Correlation | | | | | |
| US Equities | 12.2% | 17.4% | 1.00 | | | | | |
| Foreign Equities | 10.8% | 20.5% | 0.67 | 1.00 | | | | |
| Treasury Bonds | 6.7% | 7.0% | 0.15 | -0.03 | 1.00 | | | |
| Corporate Bonds | 7.5% | 9.0% | 0.51 | 0.26 | 0.78 | 1.00 | | |
| Commodities | 7.6% | 23.5% | 0.03 | 0.16 | -0.13 | -0.11 | 1.00 | |
| Cash | 4.4% | 3.6% | 0.06 | 0.08 | 0.47 | 0.22 | 0.19 | 1.00 |

Exhibit 3 shows two optimal portfolios using mean-variance analysis and the inputs in the top two panels of Exhibit 2 for a low growth regime. Both approaches were constrained to have expected returns of 8.0%. The key takeaway from Exhibit 3 is that both approaches yield roughly the same overall exposure to equity and fixed income assets.

Exhibit 3: Regime Conditioned Optimal Portfolios

| | Relevance Approach | Conventional Approach |
|---|---|---|
| US equities | 0.0% | 0.0% |
| Foreign Equities | 14.2% | 5.5% |
| US Treasuries | 57.2% | 65.3% |
| Corporate Bonds | 0.0% | 1.6% |
| Commodities | 0.0% | 0.0% |
| Risk Free Asset | 28.6% | 27.6% |
| Expected Return | 8.0% | 8.0% |
| Standard Deviation | 5.7% | 5.5% |

Next, we consider regime sensitive portfolios in which the regimes are defined by two indicators: inflation and real growth. We specify four regimes using these indicators.

| | | |
|---|---|---|
| Robust | Inflation = 0.9% | Real Growth = 4.5% |
| Overheated | Inflation = 5.0% | Real Growth = 4.5% |
| Downturn | Inflation = 0.9% | Real Growth = 1.9% |
| Stagflation | Inflation = 5.0% | Real Growth = 0.9% |

Exhibit 4 shows a scatter plot of inflation and real growth (in excess of their respective averages) for our full sample of years from 1974 through 2022.

Exhibit 4: Historical Inflation and Real Growth Relative to Average

1974 through 2022

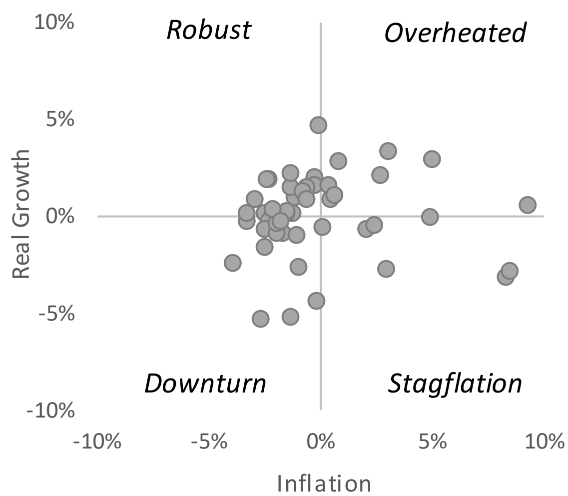


Exhibit 5 shows the relevance of each period given these two indicators for each regime. Notice that that these historical observations are not exclusively relevant to any one regime. The year 1984, for example, is relevant for estimating asset class behavior for both a Robust regime as well as an Overheated regime. This feature of our approach relates to the desirable ambiguity of regime identification that we discussed earlier. Given the available information, it might not be possible to assign every period to a unique regime with perfect confidence. Our approach addresses this uncertainty in a rigorous way.

Exhibit 5: Multivariate Relevance of Historical Observations based on Inflation and Real Growth

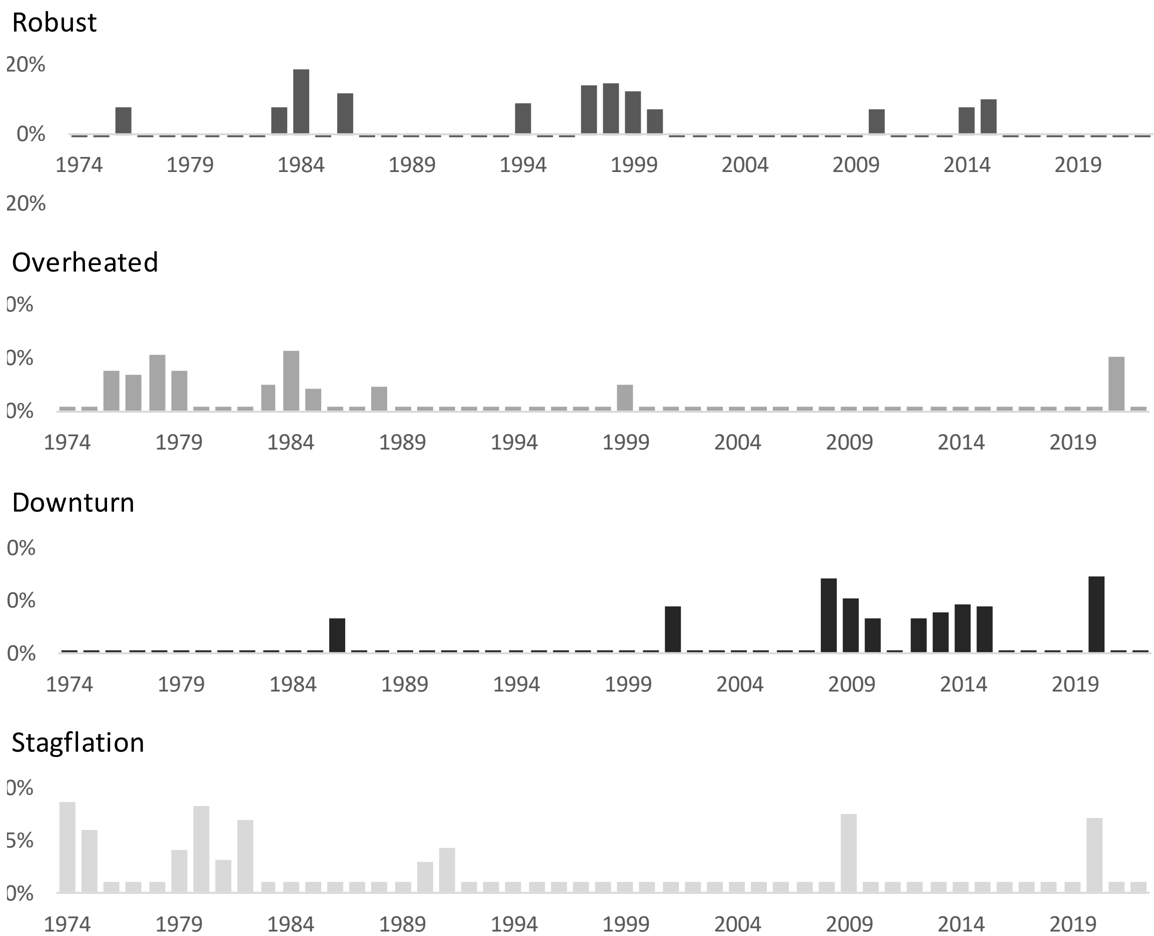


Exhibit 6 shows mean-variance optimal portfolio allocations based on the relevance-based conditional expected returns and covariances for each regime.[3] The allocations differ substantially across regimes, and though some portfolios contain zero allocation to particular asset classes, every asset class is represented in at least two of the portfolios. The allocations reveal some interesting and intuitive results:

- The portfolios for robust and overheated regimes contain larger equity allocations than the portfolios for downturn and stagflation regimes.
- International diversification is more prominent for the overheated and stagflation regimes than for the others.
- US Treasuries are featured heavily in the downturn regime portfolio, but the stagflation regime portfolio favors incremental exposure to commodities and the risk free asset.
- Corporate bonds receive the largest allocation during robust regimes, and none during stagflation.

The ambiguity of regimes also implies that investors will not have perfect confidence in their prediction of which regime will prevail in the upcoming investment period. Regime conditioned portfolios can be combined in proportion to the prospective probability an investor assigns to each regime. Alternatively, one may prefer to construct a mean-variance optimal portfolio that reflects blended inputs for expected returns and covariances, as shown in the final column of Exhibit 6. We use traditional mean-variance optimization in this example, but due to the transparent nature of the relevance methodology it is also possible to account for complexities in the return distributions of assets such as skewness, kurtosis, or asymmetric correlations, as well as customized utility preferences applied to each outcome in the historical sample.

Exhibit 6: Regime Conditioned Portfolios for Multivariate Scenarios

| **Scenario Definitions** | Robust | Overheated | Downturn | Stagflation | Blend* |
|---|---|---|---|---|---|
| Inflation Assumption | 0.9% | 5.0% | 0.9% | 5.0% | |
| Real Growth Assumption | 4.5% | 4.5% | 1.9% | 0.9% | |
| **Allocations** | | | | | |
| US Equities | 56.7% | 26.4% | 9.3% | 15.1% | 46.5% |
| Foreign Equities | 0.4% | 26.8% | 0.0% | 5.8% | 0.0% |
| US Treasuries | 0.0% | 0.0% | 48.9% | 36.1% | 53.5% |
| Corporate Bonds | 35.8% | 22.3% | 15.1% | 0.0% | 0.0% |
| Commodities | 7.1% | 24.4% | 0.0% | 9.7% | 0.0% |
| Risk Free Asset | 0.0% | 0.0% | 26.6% | 33.2% | 0.0% |
| **Conditional Performance** | | | | | |
| Expected Return | 12.4% | 13.3% | 4.8% | 8.0% | 8.8% |
| Standard Deviation | 3.8% | 9.4% | 3.4% | 5.4% | 9.1% |

* The blended portfolio uses a weighted average of expected returns, standard deviations, and correlations give weights of 10% Robust, 30% Overheated, 40% Downturn, and 20% Stagflation.

## Conclusion

Investors typically build regime specific portfolios by observing past periods in which such a regime prevailed based on a single indicator and a fixed threshold, and by computing estimates of expected returns, standard deviations, and correlations by averaging their values during these periods. This conventional approach of equally weighting observations from past regimes assumes that a regime either occurred or did not occur unambiguously and that each period is equally relevant to assessing asset class behavior in a forthcoming regime whether the regime indicator was only marginally beyond the regime threshold or far beyond the threshold. This binary view of a regime is also problematic because there is no non-arbitrary way to combine more than a single regime indicator to define a regime and because it precludes the anticipation of regimes that have not occurred historically.

We propose that investors instead use a statistic called relevance to estimate asset class expected returns, standard deviations, and correlations.  Relevance measures the importance of an observation to forming an estimate in a mathematically precise way.  It is composed of two components, similarity and informativeness, both of which are measured as squared z-scores for regimes that are measured by a single indicator, or as Mahalanobis distances in the case of regimes that are measured by more than a single indicator.  This relevance-based approach to estimating regime characteristics recognizes that regimes do not occur in an unambiguous, binary fashion; rather, there are degrees to which a regime prevails.  This non-binary description of regimes also enables us to define regimes based on the co-occurrence of multiple indicators in a way that is theoretically justified, and it allows us to contemplate regimes that have not occurred historically but are nonetheless plausible looking forward.

**Appendix**

Relevance-weighted average of prior outcomes equals linear regression prediction

The prediction equation corresponding to full sample linear regression equals:

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

Expanding out the expression for relevance gives:

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

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

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

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

Where:

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

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

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

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

$$\hat{y}_t = \alpha + x_t\beta \tag{A8}$$

## 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

M. Czasonis, M. Kritzman, and D. Turkington. 2022. *Prediction Revisited: The Importance of Observation*, John S. Wiley & Sons. pp. 155 - 162.

M. Czasonis, M. Kritzman, and D. Turkington. 2023. “Relevance-Based Prediction: A Transparent and Adaptive Alternative to Machine Learning.” *The Journal of Financial Data Science*, Winter, vol. 5, no. 1.

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] Shannon showed that information is an inverse logarithmic function of probability, which is a key insight from his comprehensive theory of communication. See Shannon (1948)

[2] Rather than choose a threshold for relevance arbitrarily or heuristically, we can use a metric called fit to determine the optimal subsample size. Fit measures the alignment of relevance and outcomes. See Czasonis, Kritzman, and Turkington (2023) for more detail about this approach.

[3] We derived these portfolios by first estimating the return that would be expected in each regime from a 60/40 stock/bond portfolio. We then applied mean-variance analysis using the regime-specific relevance weighted estimates of expected returns, standard deviations, and correlations to minimize risk given these regime dependent expected returns for a 60/40 portfolio.