A Linear Regression's Predictions are Relevance-Weighted Averages of Past Outcomes

Yannick Kälber · 7 Feb 2026

https://functor.network/user/3370/entry/1645

The book *Asset Allocation: From Theory to Practice and Beyond* by Kinlaw et al. (2021) is one of my favorites as it gets a few myths about mean variance optimization right, which are constantly parroted, even in academic papers to motivate some fancy new method as solution. It also provides useful solutions to a portfolio manager's or allocator's practical questio like when to rebalance.

In chapter 13 on *Forecasting* the authors introduce a method called **partial sample regression / relevance-based prediction**, which provides an interesting perspective on a linear regression's prediction and generalizes it into a direction akin to the *Nadaraya-Watson* kernel regression.

Linear Regression

Say we are in a time-series setting with N past observations of input variables $\mathbf{x}_i \in \mathbb{R}^{1\times K}$ and a corresponding output variable $y_{i+1} \in \mathbb{R}$. Now, at $t > i$, we get realizations of $\mathbf{x}_t$ and are asked to predict y_{t+1}.

What's the prediction of a linear regression? We start with the general form and rewrite it in terms of de-meaned quantities. In the following, $\boldsymbol{\beta} \in \mathbb{R}^{(K+1)\times 1}$ is the full coefficient vector including the intercept and $\tilde{\boldsymbol{\beta}} = (\beta_1, \beta_2, \dots, \beta_K)' \in \mathbb{R}^{K\times 1}$ is the vector of slope coefficients only.

$$
\begin{aligned}
\hat{y}_{t+1} &= (1,\, \mathbf{x}_t)\boldsymbol{\beta} && (1)\\
&= \mathbf{x}_t\tilde{\boldsymbol{\beta}} + \beta_0 && (2)\\
&= \mathbf{x}_t\tilde{\boldsymbol{\beta}} + (\bar{y} - \bar{\mathbf{x}}\tilde{\boldsymbol{\beta}}) && (3)\\
&= (\mathbf{x}_t - \bar{\mathbf{x}})\tilde{\boldsymbol{\beta}} + \bar{y} && (4)\\
&= \tilde{\mathbf{x}}_t\tilde{\boldsymbol{\beta}} + \bar{y} && (5)
\end{aligned}
$$

In (2) we separate the intercept $\beta_0 \in \mathbb{R}$ from the slope coefficients, noting that $(1,\, \mathbf{x}_t)\boldsymbol{\beta} = \beta_0 + \mathbf{x}_t\tilde{\boldsymbol{\beta}}$. In (3) we substitute the least squares expression for the intercept, $\beta_0 = \bar{y} - \bar{\mathbf{x}}\tilde{\boldsymbol{\beta}}$. In (4) we collect the terms that multiply $\tilde{\boldsymbol{\beta}}$. In (5) we define $\tilde{\mathbf{x}}_t = \mathbf{x}_t - \bar{\mathbf{x}} \in \mathbb{R}^{1\times K}$ as the de-meaned current observation. The sample means are:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \in \mathbb{R}^{1\times K} \quad \text{and} \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_{i+1} \in \mathbb{R}$$

Tedious Algebra

The reason I reformulated the linear regression's prediction a bit is to set the stage for further reformulations. Equation (5) shows that the prediction is the sample mean of the outcome plus a correction term. To work with the correction term in isolation, we subtract $\bar{y}$ from both sides of (5):

$$\tilde{\hat{y}}_{t+1} = \hat{y}_{t+1} - \bar{y} = \tilde{\mathbf{x}}_t\tilde{\boldsymbol{\beta}} \tag{6}$$

Now we substitute the least squares estimator $\tilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i\right)^{-1}\left(\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{y}_{i+1}\right)$ into (6) and rearrange:

$$\begin{aligned}
\tilde{\hat{y}}_{t+1} &= \tilde{\mathbf{x}}_t\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{y}_{i+1}\right) && (7)\\
&= \tilde{\mathbf{x}}_t\left(\frac{1}{N-1}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i\right)^{-1}\left(\frac{1}{N-1}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{y}_{i+1}\right) && (8)\\
&= \tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\left(\frac{1}{N-1}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{y}_{i+1}\right) && (9)\\
&= \frac{1}{N-1}\sum_{i=1}^{N}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i'\,\tilde{y}_{i+1} && (10)
\end{aligned}$$

In (7) we divide both the matrix in the inverse and the vector on the right by N. This is permissible because $\frac{1}{N}$ cancels between numerator and denominator. In (8) we replace $\frac{1}{N}$ with $\frac{1}{N-1}$ in both factors, which again cancels, but $\frac{1}{N-1}$ is the conventional normalisation for sample covariance matrices. In (9) we recognise the sample covariance matrix $\boldsymbol{\Omega} = \frac{1}{N-1}\sum_{i=1}^{N}\tilde{\mathbf{x}}_i'\tilde{\mathbf{x}}_i \in \mathbb{R}^{K\times K}$. In (10) we pull $\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}$ inside the sum, since it does not depend on the summation index i.

Equation (10) is already revealing: the de-meaned prediction is a weighted sum of de-meaned outcomes, where the weight on observation i is the inner product $\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i'$. But we can go further by decomposing this inner product into geometrically interpretable pieces.

Because $\tilde{y}_{i+1} = y_{i+1} - \bar{y}$ is de-meaned, its sum vanishes: $\sum_{i=1}^{N} \tilde{y}_{i+1} = 0$. Any constant multiplied by this sum is therefore zero. In particular, $\frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t'$ is a scalar that does not depend on i. Subtracting it inside the sum does not change the result:

$$\tilde{\hat{y}}_{t+1} = \frac{1}{N-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \, \tilde{y}_{i+1} \tag{11}$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \, \tilde{y}_{i+1} - \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' \cdot \underbrace{\frac{1}{N-1} \sum_{i=1}^{N} \tilde{y}_{i+1}}_{=\,0} \tag{12}$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} \left(\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' - \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' \right) \tilde{y}_{i+1} \tag{13}$$

(11) repeats (10). In (12) we subtract a term that equals zero. In (13) we absorb the constant into the sum.

We now want to replace the bilinear form $\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i'$ inside (13) with something more interpretable. Since $\mathbf{\Omega}^{-1}$ is symmetric and positive definite, $\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{\Omega}^{-1}} = \mathbf{a}\mathbf{\Omega}^{-1}\mathbf{b}'$ defines an inner product, and $\|\mathbf{a}\|^2_{\mathbf{\Omega}^{-1}} = \mathbf{a}\mathbf{\Omega}^{-1}\mathbf{a}'$ its induced (squared) norm. Every inner product satisfies the **polarisation identity**, which we verify by direct expansion:

$$\frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' + \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' - \frac{1}{2}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t) \mathbf{\Omega}^{-1} (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)' \tag{14}$$

$$= \frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' + \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' - \frac{1}{2}\left(\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' - \tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' - \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' + \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' \right) \tag{15}$$

$$= \frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' + \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' - \frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' + \frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' + \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' - \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' \tag{16}$$

$$= \frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' + \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \tag{17}$$

$$= \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \tag{18}$$

In (15) we expand the quadratic form $(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)\mathbf{\Omega}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)'$. In (16) we distribute the $-\frac{1}{2}$. In (17) the squared-norm terms cancel pairwise: $+\frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i'$ with $-\frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i'$, and $+\frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t'$ with $-\frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t'$. In (18) we use the symmetry of $\mathbf{\Omega}^{-1}$, which implies $\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' = \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i'$.

Reading (18) from right to left gives us the decomposition we need:

$$\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' = \frac{1}{2}\tilde{\mathbf{x}}_i \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' + \frac{1}{2}\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_t' - \frac{1}{2}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t) \mathbf{\Omega}^{-1} (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)' \tag{19}$$

This is the matrix-weighted analogue of the scalar identity $ab = \frac{1}{2}(a^2 + b^2 - (a-b)^2)$. We substitute (19) into (13):

$$\tilde{\hat{y}}_{t+1} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' - \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'\right)\tilde{y}_{i+1} \tag{20}$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{1}{2}\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' + \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t' - \frac{1}{2}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)' - \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega} \tag{21}$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{1}{2}\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' - \frac{1}{2}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)'\right)\tilde{y}_{i+1} \tag{22}$$

(20) repeats (13). In (21) we replace $\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i'$ using (19). In (22) the terms $+\frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'$ from (19) and $-\frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'$ from (20) cancel.

We add $\bar{y}$ to both sides of (22) to recover the actual prediction. Since $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, we also restore the original (un-centered) notation inside the quadratic forms, noting that $\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' = (\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\Omega}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})'$ and $(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t) = (\mathbf{x}_i - \mathbf{x}_t)$:

$$\hat{y}_{t+1} = \bar{y} + \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\Omega}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})' - \frac{1}{2}(\mathbf{x}_i - \mathbf{x}_t)\boldsymbol{\Omega}^{-1}(\mathbf{x}_i - \mathbf{x}_t)'\right)(y \tag{23}$$

Predictor Space Geometry

Let's inspect the two quadratic forms in (23). Both are instances of the **Mahalanobis distance**, a distance measure that accounts for the variance of and covariance between the input variables through the inverse covariance matrix $\boldsymbol{\Omega}^{-1}$. Where the ordinary Euclidean distance treats every direction in predictor space equally, the Mahalanobis distance stretches directions of low variance and compresses directions of high variance, effectively standardizing the measurement. In the special case $\boldsymbol{\Omega} = \mathbf{I}$ (uncorrelated variables with unit variance), the Mahalanobis distance reduces to the Euclidean distance.

The first term, $\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\Omega}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})'$, is (half) the squared Mahalanobis distance of observation i from the historical average. It measures how *unusual* this observation was relative to typical conditions. Kinlaw et al. call this quantity **Informativeness** $I(\mathbf{x}_i)$: observations far from the mean are more likely to carry real event-driven information, while near-average observations are more likely to be noise.

The second term, $\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_t)\mathbf{\Omega}^{-1}(\mathbf{x}_i - \mathbf{x}_t)'$, is (half) the squared Mahalanobis distance between observation i and the *current* conditions $\mathbf{x}_t$. Its negative, which enters with a minus sign, is what Kinlaw et al. call **Similarity** $S(\mathbf{x}_i, \mathbf{x}_t)$: observations close to the current situation are similar, and the negative Mahalanobis distance is largest (closest to zero) for observations that resemble the present.

With these definitions in hand, we reorder the terms in (23) and label them:

$$\hat{y}_{t+1} = \bar{y} + \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\underbrace{-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_t)\mathbf{\Omega}^{-1}(\mathbf{x}_i - \mathbf{x}_t)'}_{S(\mathbf{x}_i,\,\mathbf{x}_t)} + \underbrace{\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}})\mathbf{\Omega}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})'}_{I(\mathbf{x}_i)}\right)\right] \tag{24}$$

Kinlaw et al. (2021) combine these two components into a single metric called **Relevance**:

$$R(\mathbf{x}_i, \mathbf{x}_t) = S(\mathbf{x}_i, \mathbf{x}_t) + I(\mathbf{x}_i) \tag{25}$$

Substituting (25) into (24):

$$\hat{y}_{t+1} = \bar{y} + \frac{1}{N-1}\sum_{i=1}^{N}\left[R(\mathbf{x}_i, \mathbf{x}_t)\,(y_{i+1} - \bar{y})\right] \tag{26}$$

Relation to the Hat Matrix

The relevance decomposition is closely related to the classical **hat matrix** from regression theory. The hat matrix $\mathbf{H} \in \mathbb{R}^{N\times N}$ is defined as the matrix that maps the vector of observed outcomes $\mathbf{y} \in \mathbb{R}^N$ to the vector of fitted values $\hat{\mathbf{y}} \in \mathbb{R}^N$:

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}, \qquad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{27}$$

Here $\mathbf{X} \in \mathbb{R}^{N\times(K+1)}$ is the design matrix whose rows are $\mathbf{x}_i^* = (1,\, \mathbf{x}_i) \in \mathbb{R}^{1\times(K+1)}$ (augmented with a leading 1 for the intercept). The element H_{ji} determines how much the outcome of observation i contributes to the fitted value at observation j. For an out-of-sample prediction at $\mathbf{x}_t$, the same logic applies: $\hat{y}_{t+1} = \mathbf{x}_t^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \sum_{i=1}^{N} h_{ti}\, y_{i+1}$, where $h_{ti} = \mathbf{x}_t^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i^{*\prime}$ is the weight that observation i receives. The name “hat matrix” comes from the fact that it puts the hat on $\mathbf{y}$.

We can derive the hat matrix weights directly from our existing equations. Starting from (10), we substitute $\tilde{y}_{i+1} = y_{i+1} - \bar{y}$ and expand:

$$\hat{y}_{t+1} = \bar{y} + \frac{1}{N-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \, \tilde{y}_{i+1} \tag{28}$$

$$= \frac{1}{N} \sum_{i=1}^{N} y_{i+1} + \frac{1}{N-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \, (y_{i+1} - \bar{y}) \tag{29}$$

$$= \frac{1}{N} \sum_{i=1}^{N} y_{i+1} + \frac{1}{N-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \, y_{i+1} - \frac{\bar{y}}{N-1} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_i' \tag{30}$$

$$= \frac{1}{N} \sum_{i=1}^{N} y_{i+1} + \frac{1}{N-1} \sum_{i=1}^{N} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \, y_{i+1} \tag{31}$$

$$= \sum_{i=1}^{N} \left(\frac{1}{N} + \frac{1}{N-1} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \right) y_{i+1} \tag{32}$$

$$= \sum_{i=1}^{N} h_{ti} \, y_{i+1} \tag{33}$$

In (29) we write $\bar{y} = \frac{1}{N} \sum y_{i+1}$ and substitute $\tilde{y}_{i+1} = y_{i+1} - \bar{y}$. In (30) we distribute the product and pull $\bar{y}$ out of the second sum. In (31) we use $\sum_{i=1}^{N} \tilde{\mathbf{x}}_i' = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})' = \mathbf{0}$, which eliminates the last term. In (32) we collect both sums into a single sum over y_{i+1}.

Equation (33) reveals the structural similarity to the **Nadaraya-Watson** kernel regression. In both cases, the prediction is simply a weighted sum of past outcomes. The difference lies in how the weights h_{ti} are constructed. Reading off the coefficient of y_{i+1} in (33), the hat matrix weight is:

$$h_{ti} = \frac{1}{N} + \frac{1}{N-1} \tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i' \tag{34}$$

The first part, $\frac{1}{N}$, is a uniform contribution: every observation participates equally in the sample mean. The second part is the inner product $\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i'$, scaled by $\frac{1}{N-1}$, which tilts the weight toward or away from observation i depending on the geometric relationship between $\mathbf{x}_i$ and $\mathbf{x}_t$.

How does the relevance $R(\mathbf{x}_i, \mathbf{x}_t)$ relate to the inner product $\tilde{\mathbf{x}}_t \mathbf{\Omega}^{-1} \tilde{\mathbf{x}}_i'$? Let us expand $S + I$ explicitly:

$$
\begin{aligned}
R(\mathbf{x}_i, \mathbf{x}_t) &= S(\mathbf{x}_i, \mathbf{x}_t) + I(\mathbf{x}_i) && (35)\\
&= -\frac{1}{2}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)' + \frac{1}{2}\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' && (36)\\
&= -\frac{1}{2}\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' + \tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' - \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t' + \frac{1}{2}\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' && (37)\\
&= \tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' - \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t' && (38)
\end{aligned}
$$

In (37) we expand $(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_t)'$ and use the symmetry of $\boldsymbol{\Omega}^{-1}$. In (38) the $\pm\frac{1}{2}\tilde{\mathbf{x}}_i\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i'$ terms cancel.

Relevance and the inner product $\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i'$ differ by the residual term $-\frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'$. This scalar depends on the current conditions $\mathbf{x}_t$ but not on the summation index i. In the prediction (26), it is multiplied by $\sum_{i=1}^{N}(y_{i+1} - \bar{y}) = 0$ and therefore vanishes. The hat matrix weight and relevance are thus related by:

$$
\begin{aligned}
h_{ti} &= \frac{1}{N} + \frac{1}{N-1}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' && (39)\\
&= \frac{1}{N} + \frac{1}{N-1}\left(R(\mathbf{x}_i, \mathbf{x}_t) + \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'\right) && (40)\\
&= \underbrace{\frac{1}{N} + \frac{\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'}{2(N-1)}}_{c_t} + \frac{R(\mathbf{x}_i, \mathbf{x}_t)}{N-1} && (41)
\end{aligned}
$$

In (40) we substitute (38) solved for the inner product: $\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_i' = R(\mathbf{x}_i, \mathbf{x}_t) + \frac{1}{2}\tilde{\mathbf{x}}_t\boldsymbol{\Omega}^{-1}\tilde{\mathbf{x}}_t'$. In (41) we collect the terms that do not depend on i into the constant c_t.

The term c_t depends only on the current conditions $\mathbf{x}_t$ and acts as a uniform baseline for every observation. It combines the $\frac{1}{N}$ from the sample mean with a correction that accounts for how far $\mathbf{x}_t$ lies from the historical centre. The observation-specific part of h_{ti} is proportional to relevance alone.

This means the hat matrix and the relevance representation agree on the *prediction* (both yield $\hat{y}_{t+1}$), and the per-observation variation in the hat matrix weights is governed entirely by $R(\mathbf{x}_i, \mathbf{x}_t)$. The relevance decomposition thus provides a geometric interpretation of what the hat matrix encodes: similarity to the current conditions and informativeness of the historical observation.

What Equation (26) Reveals

Equation (26) is an interesting rewriting of the linear regression prediction $\hat{y}_{t+1} = (1, \mathbf{x}_t)\boldsymbol{\beta}$. It produces the same number for any dataset and any $\mathbf{x}_t$. But it reveals the mechanics that the coefficient-based formula conceals.

The prediction is anchored at the historical average $\bar{y}$ and then adjusted by a weighted sum of past outcome deviations $y_{i+1} - \bar{y}$. The weight on each historical observation i is its relevance $R(\mathbf{x}_i, \mathbf{x}_t)$, which combines two Mahalanobis-distance-based quantities: how similar $\mathbf{x}_i$ is to the current conditions $\mathbf{x}_t$, and how informative $\mathbf{x}_i$ was to begin with. As (41) shows, relevance is the observation-varying component of the hat matrix weight; the hat matrix and the relevance representation describe the same linear combination of outcomes from two complementary perspectives.

Observations with positive relevance pull the prediction toward their outcomes in the natural way. Observations with *negative* relevance, however, contribute too, and they contribute in reverse. The regression uses the outcome of a dissimilar, uninformative period and assumes that the opposite will occur. It treats this inversion as equally valuable as the direct extrapolation from relevant periods. However, when the data contains regime shifts, the outcomes of an irrelevant period carry no valuable information about the current one. Inverting them is expected to worsen the forecast.

Equation (26) therefore sets the stage for a natural generalization. If the use of negatively relevant observations is problematic, the remedy is to simply exclude them: restrict the complete set of N observations to a partial subset of $n \leq N$ with $R(\mathbf{x}_i, \mathbf{x}_t) > R^*$ (R^* is a relevance threshold, for instance 0). This is what Kinlaw et al. (2021) call **Partial Sample Regression**.

References

Kinlaw, W., Kritzman, M., and Turkington, D. (2021): *Asset Allocation: From Theory to Practice and Beyond*. Wiley.