

Calculating the Variance of Weighted Results to Optimize Weighting

Adam Dorhauer

In my article "The Math of Weighting Past Results" on the Hardball Times, I gave a formula for finding the proper weighting for past data given certain inputs from the dataset. This formula defined the relationship between weighted results and talent, and the proper weighting was the value that maximized that relationship.

I started with a formula for a sample with exactly two days and then generalized that to cover any length of sample. I more or less explained where the two-day version came from in the article, but not the full version, which was as follows:

Expected Correlation Between Weighted Results and Current Talent for Any Length Sample

w: decay factor for weighting past results

r: day-to-day correlation of talent

d: number of days in the sample

Var_{true} : variance in current talent levels

Var_{x1} : observed variance in results over one day

$$sd_x = \sqrt{2\text{Var}_{\text{true}} \left(\frac{(rw - (rw)^d)}{(1 - rw)(1 - w^2)} + \frac{w^{2d} \left(\left(\frac{r}{w}\right)^d - rw \right)}{\left(1 - \frac{r}{w}\right)(1 - w^2)} \right) + \text{Var}_{x1} \left(\frac{1 - w^{2d}}{1 - w^2} \right)}$$

$$sd_y = \sqrt{\text{Var}_{\text{true}}}$$

$$\text{Cov}_{x,y} = \text{Var}_{\text{true}} \left(\frac{1 - (rw)^d}{1 - rw} \right)$$

$$\textit{expected correlation} = \frac{\text{Cov}_{x,y}}{sd_x sd_y}$$

This supplement will go through the calculations of generalizing the simpler two-day formula to what we see above. It will rely heavily on the use of geometric series, so I would recommend having some familiarity with those before attempting to follow these calculations.

In the article, we treated each day's results as a separate variable and the overall sample as a sum of these individual daily variables. When we had two days in our sample, the combined variance was defined by the following formula:

$$Var_{(x_1+x_2)} = Var_{x_1} + Var_{x_2} + 2Cov_{x_1,x_2}$$

This formula can be expanded to include more than two variables, but it starts to get messy really quickly. To make expanding simpler, the formula can be re-written as a covariance matrix. If you have n variables, then the covariance matrix will be an $n \times n$ array, where each entry is the covariance between the variables representing that row and column. For two days, we would fill in the covariance matrix as follows:

	x_1	x_2
x_1	Var_{x_1}	Cov_{x_1,x_2}
x_2	Cov_{x_1,x_2}	Var_{x_2}

The combined variance is equal to the sum of the items in the matrix, which you can see is equivalent to the above formula.

This makes it much simpler to expand the formula for additional variables since you just have to add more rows and columns to the matrix. In the article, we found that the day-to-day correlation of talent (r) and the decay factor used to weight past data (w) can be used to explain changes in the variances and covariances throughout the sample:

$$Var_{(wx_1+x_2)} = w^2 * Var_{x_1} + Var_{x_1} + 2rw * Var_{true}$$

Translating this to our covariance matrix gives us (with Var_{x_1} and Var_{true} written as v_x and v_t to save space):

	x_1	wx_2
x_1	v_x	rw^*v_t
wx_2	rw^*v_t	w^2*v_x

If we expand this to include additional days, every term except those on the diagonal will include a Var_{true} factor, and those on the diagonal will instead have a Var_{x_1} factor. (This is because the terms on the diagonal represent the covariance of each

variable with itself, which is just the variance of that variable.) Similarly, every term contains an r factor and a w factor, except that the terms on the diagonal have no r (because these are relating the results of one day to themselves, so it is irrelevant how much talent changes from day to day).

For now, let's strip out the variance factors and focus only on what happens to r and w as we expand the matrix to cover more days. We'll look at r and w separately, but keep in mind these are just factors from the same matrix, not two separate matrices. If you placed one on top of the other, so that each r term lines up with the corresponding w term, and then put the variances back in, you'd get the full matrix.

This covariance matrix is essentially the same as what we worked with for the variance article, except now we are introducing weights for past results. As a result, the only real difference here is what happens with the w 's, and the r terms follow the same pattern as in math for the variance article:

	x_1	wx_2	w^2x_3	w^3x_4	...	$w^{d-1}x_d$
x_1	r^0	r^1	r^2	r^3	...	r^{d-1}
wx_2	r^1	r^0	r^1	r^2	...	r^{d-2}
w^2x_3	r^2	r^1	r^0	r^1	...	r^{d-3}
w^3x_4	r^3	r^2	r^1	r^0	...	r^{d-4}
:	:	:	:	:	⋮	⋮
$w^{d-1}x_d$	r^{d-1}	r^{d-2}	r^{d-3}	r^{d-4}	...	r^0

The weights also follow a pattern, though not the same one as the r factors. The weight for each term equals the combined weight of the two variables it represents:

	x_1	wx_2	w^2x_3	w^3x_4	...	$w^{d-1}x_d$
x_1	w^0	w^1	w^2	w^3	...	w^{d-1}
wx_2	w^1	w^2	w^3	w^4	...	w^d
w^2x_3	w^2	w^3	w^4	w^5	...	w^{d+1}
w^3x_4	w^3	w^4	w^5	w^6	...	w^{d+2}
:	:	:	:	:	⋮	⋮
$w^{d-1}x_d$	w^{d-1}	w^d	w^{d+1}	w^{d+2}	...	$w^{2(d-1)}$

While the two patterns are different, there are three important things to note that hold for both of them:

- 1) The terms on the main diagonal form their own distinct pattern.
- 2) The remaining terms are symmetrical about the diagonal, with the terms above and below the diagonal mirroring each other.
- 3) The terms on each diagonal parallel to the main diagonal follow a distinct pattern.

We need to find the sum of the matrix to get the variance in the weighted results. Using these three observations, we can simplify the sum by dividing the matrix up into parts.

We'll start with the main diagonal of the matrix. The terms on the diagonal follow the form $w^{2i}v_x$. The sum of these terms is a geometric series, which makes it simple to evaluate:

$$\sum_{i=0}^{d-1} w^{2i} Var_{x1} = Var_{x1} \left(\frac{1 - w^{2d}}{1 - w^2} \right)$$

Next, because the matrix is symmetrical about the diagonal, we can focus on the sum for only the terms above or below the diagonal and then double our result later.

We'll compute this sum by continuing to divide the matrix along its diagonal rows. The r values within a given diagonal are all identical, which we can see in this graphic from the math for the previous article on variance:

	t_0	t_1	t_2	t_3	...
t_0	r^0	r^1	r^2	r^3	...
t_1	r^1	r^0	r^1	r^2	...
t_2	r^2	r^1	r^0	r^1	...
t_3	r^3	r^2	r^1	r^0	...
:	:	:	:	:	⋮

The w values within each diagonal also follow a set pattern, though slightly more complex than the one for r 's. Rather than $r^1+r^1+r^1+\dots$, we get $w^1+w^3+w^5+\dots$. The basic pattern for the first diagonal is:

$$\sum_{i=0}^{d-2} w^{2i+1}$$

That's just for the w component of each term. If we include the r and variance components, we get this for the sum of the terms in the first diagonal adjacent the main diagonal:

$$rw^{2i+1}Var_{true}$$

$$rVar_{true} \sum_{i=0}^{d-2} w^{2i+1}$$

$$\begin{aligned}
rwVar_{true} \sum_{i=0}^{d-2} w^{2i} &= rwVar_{true} \left(\frac{1 - w^{2(d-1)}}{1 - w^2} \right) \\
&= rVar_{true} \left(\frac{w - w^{2d-1}}{1 - w^2} \right)
\end{aligned}$$

This is still a geometric series, so we can evaluate the sum for this diagonal.

For the second diagonal, the w's go $w^2 + w^4 + w^6 + \dots$, which gives us:

$$r^2Var_{true} \sum_{i=1}^{d-2} w^{2i} = r^2Var_{true} \left(\frac{w^2 - w^{2d-2}}{1 - w^2} \right)$$

If we keep going, we'll find that for each additional diagonal, the exponent for r will rise by one, the starting value of i in the summation will rise by one (which also means the summation will have one fewer term, which we can see by looking at the matrix), and each diagonal will alternate having an extra w outside the geometric sum due to the diagonals alternating between odd and even exponents.

Fortunately, the alternating w problem disappears when distribute that w back into the result for the geometric sum of each odd diagonal. We end up with the following pattern for the sum of each diagonal (after factoring out the Var_{true} component from each term):

$$\begin{aligned}
Var_{true} * &\left(r \left(\frac{w - w^{2d-1}}{1 - w^2} \right) + r^2 \left(\frac{w^2 - w^{2d-2}}{1 - w^2} \right) + r^3 \left(\frac{w^3 - w^{2d-3}}{1 - w^2} \right) \right. \\
&\left. + \dots r^{d-1} \left(\frac{w^{d-1} - w^{2d-(d-1)}}{1 - w^2} \right) \right)
\end{aligned}$$

$$\begin{aligned}
Var_{true} * &\left(\frac{rw - rw^{2d-1} + r^2w^2 - r^2w^{2d-2} + \dots r^{d-1}w^{d-1} - r^{d-1}w^{2d-(d-1)}}{1 - w^2} \right) \\
Var_{true} * &\left(\frac{rw + r^2w^2 + \dots r^{d-1}w^{d-1}}{1 - w^2} + \frac{-rw^{2d-1} - r^2w^{2d-2} - \dots r^{d-1}w^{2d-(d-1)}}{1 - w^2} \right)
\end{aligned}$$

$$Var_{true} * \left(\frac{rw + (rw)^2 + \dots (rw)^{d-1}}{1 - w^2} + \frac{-w^{2d} \left(\frac{r}{w} + \left(\frac{r}{w}\right)^2 + \dots \left(\frac{r}{w}\right)^{d-1} \right)}{1 - w^2} \right)$$

This gives us two separate geometric series: the first multiplies by a factor of rw , and the second by a factor of r/w . Simplifying these geometric series gives us:

$$Var_{true} \left(\frac{(rw - (rw)^d)}{(1 - rw)(1 - w^2)} + \frac{w^{2d} \left(\left(\frac{r}{w}\right)^d - rw \right)}{\left(1 - \frac{r}{w}\right)(1 - w^2)} \right)$$

That gives us the sum of everything above the main diagonal in the covariance matrix. To get the full sum of the matrix, we need to double this (to account for everything below the diagonal, which mirrors this calculation) and add the sum of the main diagonal:

$$2Var_{true} \left(\frac{(rw - (rw)^d)}{(1 - rw)(1 - w^2)} + \frac{w^{2d} \left(\left(\frac{r}{w}\right)^d - rw \right)}{\left(1 - \frac{r}{w}\right)(1 - w^2)} \right) + Var_{x1} \left(\frac{1 - w^{2d}}{1 - w^2} \right)$$

This gives us the full variance of the weighted results. Our formula calls for the standard deviation instead of the variance, so we just take the square root of this.

Next, we need to calculate the covariance between current talent and the weighted observations. We can get this using another covariance matrix based on the idea of "shared" variance mentioned in the Hardball Times article. The covariance between the results and talent for a given day is the same as the variance in talent, since the variance in talent is inherent in the variance of the results (i.e. that variance is shared between the results and the talent levels for that day).

To fill out the rest of the covariance matrix, we use the fact that the covariance between results and current talent drops the further the results are from the

present time. The amount the covariance drops is determined by the day-to-day correlation in talent and the weight given to past data:

	x_1	wx_2	w^2x_3	w^3x_4	...	$w^{d-1}x_d$
t_1	$(rw)^0 v_t$	$(rw)^1 v_t$	$(rw)^2 v_t$	$(rw)^3 v_t$...	$(rw)^{d-1} v_t$

This is also a geometric series which multiplies by a factor of rw . The sum simplifies to:

$$Var_{true} \left(\frac{1 - (rw)^d}{1 - rw} \right)$$

As long as we know the values for r , w , Var_{true} and Var_{x1} , we can work out what the variance will be over any number of days, which means as long as we know r , Var_{true} and Var_{x1} , we can find the value of w which maximizes the relationship between weighted results and current talent.

Typically we would find this by taking the derivative of the formula and finding the point where the derivative equals 0, but given that this is a rather unpleasant derivative to calculate (and most likely will have difficult-to-find zeroes), I would strongly recommend just using the optimize function in R or some other statistical program (the calculator on the Hardball Times uses the same method to minimize/maximize a function as the optimize function in R).

One final note: this all relies on the assumption of exponential decay weighting. Exponential decay is not necessarily implied by the underlying mathematical processes; it's an assumption we are making to make our lives easier. Theoretically, we could fit the weight for each day individually, but this is far, far more complicated and not really worth the effort.

If you had 100 days in your sample, instead of maximizing the correlation for w , you would have to maximize it for a system of 100 different weight variables. If you would like to attempt this, by all means, have fun, but, while the exponential decay assumption is a simplification, it does work pretty well.

The true weight values do tend to drop slightly faster for the most recent data and then level out more for older data than exponential decay allows for, but on the whole, it doesn't make that much difference to use exponential decay.