Factor validation

We have five factors that we hope will each work 52% of the time individually, and combine to work 55% of the time. Of course we track the standalone performance of each indicator, but this is not enough. We want to monitor the theory behind each indicator using different data. This gives us twice as much chance of noticing a change in performance, and even better, a chance to notice without actually losing money.

Power validation: rematch

A simple check on the over-reaction theory is to compare the lines when two teams face each other twice in a season. If our theory is correct, whichever team is favored by more in the second game than the first (or disfavored by less, or favored in the second game but disfavored in the first game) should lose more than half the time. Since our power indicator uses only games in successive weeks, the validation is done with non-overlapping data.

The validation measure is noisier than the indicator we’re going to use for betting, because more weeks elapse between power estimation and game, and because we have fewer games to test. But it’s more direct, it does not require fitting power ratings, and it's a direct comparison of the same two teams.

If the teams play each other at the same site, it’s particularly easy, we can just compare the two lines. However this is the less common situation, most rematches are two teams in the same division playing one game in each of the two home stadiums. We could retain our assumption of a three point home field advantage to adjust the line, but I prefer to test that assumption as well. So instead I do a regression of the line change for all games of this type, with the dependent variable the amount the home team is favored in the second game and the independent variable the amount it was favored as a visitor in the first game. Positive residuals from this regression are treated as the line moving in favor of the home team in the second game, negative residuals are treated the opposite. This regression does not bias the in-sample results, because it uses no information about which team covered the spread, only the information in the line move itself.

Turnover validation: season statistics

To validate the turnover indicator, we'll look at how season turnover differential predicts wins and covers. This has minimal overlap with our turnover factor, because that looks only a single game back. Also the factor looks only at the sign of the turnover differential, the regression will use the magnitude as well.

For each game we'll look at the net turnovers for each team in the prior games for the season and subsequent games. That is, we'll add up all the times the team gave the ball away through fumble lost or interception, and subtract all the times the team took the ball away. A positive number is bad for the team, it means it lost the ball by turnover more than it gained the ball (but that's good for betting on the team the following week). Then we'll subtract the visiting team's number from the home teams.

Common sense predicts a negative relation between the home-minus-away turnover number and the home team winning the game. A positive number means that the home team had more net turnover losses, or fewer net turnover gains, than the visiting team; thus it is probably a worse team; thus it will probably lose. The relation should be symmetrical, past turnovers and future turnovers should be equally good predictors of team quality.

If our turnover indicator theory is correct, we will see the opposite relation for prior turnovers if we regress an indicator for the home team covering the spread instead of the home team winning. Although the team that gave the ball away more net is probably the worse team, we think the line overcorrects for that by ignoring the large random component to turnovers. So we expect the team with the worse turnover record to be more likely to cover.

On the other hand, the line cannot take into account subsequent turnover differential, so we would expect that to have the same negative sign in a cover regression as in a win regression.

Hunger validation: season statistics

We're going to use a similar strategy to the turnover factor validation for hunger factor validation. If our theory is correct, for any team at any point in the season, there should be a negative relation between the net times it has covered the spread in the season to date and the number of times it will cover the spread in the remaining season.

This has a bit more data overlap with the factor than in the case of turnovers, since we're using season to date net covers in both the factor and the validation. But as with turnovers, we're using only the sign for the factor, the validation uses the magnitude. And the validation uses all remaining games in the season, not just the next game.

Validation results

The table shows seven summary statistics to evaluate whether the theory behind our factors continues to hold. The first two columns are the win-loss record among all rematches between two teams within a season played at the same stadium, betting on the team the line moved against. The next two columns show the record for the more numerous home/away rematches. If the over-reaction theory behind our power indicator is sound, we should see more wins than losses in both categories.

In the first season, 2006, the over-reaction theory worked great and the subsequent five years show roughly alternating weak and strong confirmation. But in 2012 the theory failed and it has only worked once since, in 2015. The betting indicator continues to work, and the overall results favor the theory. But if I were putting money on these results, I would be doing a lot of research on why over-reaction seems to be failing in rematches.

The numbers in the remaining three columns are the t-statistics for the validation regressions. We do not expect t-statistics over two for single-season regressions. Our faith in these factors is based on multi-season statistics plus theory. If we do get a significant t-statistic in the correct direction, it is strong confirmation of our theory. A small t-statistic in the same direction is weak confirmation. A small t-statistic in the wrong direction is weak evidence against the factor, but unless we get a long run of them, we're not going to adjust our system. A significant t-statistic in the wrong direction is a red flag requiring immediate investigation.

For the turnover validation, we want the “prior” t-statistic to be positive and the “subsequent” t-statistic to be negative, and we focus mostly on the prior minus the subsequent. We see weak confirmation every year except strong confirmation in 2008 and weak disconfirmation in 2010, 2014 and 2016; although in several years one or both t-statistic has the wrong sign. The entire live period test shows strong confirmation. Overall there seems nothing to worry about.

We want negative t-statistics for the hunger indicator, but for the entire live period we have only weak confirmation. The first four years are fine, with one strong confirmation season and one weak disconfirmation. But 2010 showed strong disconfirmation and the validation has only been successful in half the years since (although 2016, which did little else good, was a strong confirmation).

So our validation numbers are sending up distress—or at least investigation demands—for three of our five indicators. But this this is a demonstration and not a business, I did not change anything.