9-18-2008

Missing Well: Optimal Targeting of Soccer Shots

Fredrick E. Vars
University of Alabama - School of Law, fvars@law.ua.edu

Missing Well: Optimal Targeting of Soccer Shots

Fredrick E. Vars

Working Paper

Missing Well: Optimal Targeting of Soccer Shots

Fredrick E. Vars

"Shoot for a post," a youth soccer coach once instructed me. It seemed like good advice at the time---the goalkeeper was in the middle and I didn't want to hit it to him---but a moment's reflection shows that my coach was wrong. If the ball went exactly where I aimed every time, I'd have hit lots of woodwork but scored few goals. In fact, I was (like all decent players) about equally likely to miss left as right, so about half of balls directed toward the goalpost would have gone wide with no chance of scoring. Better to aim at a spot on the target---but where exactly?

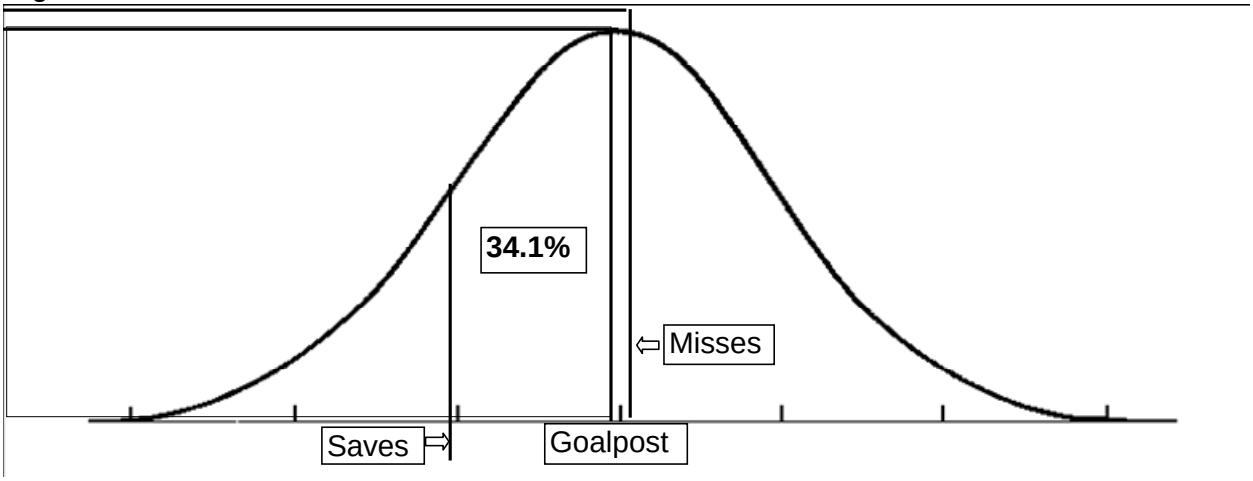
Not directly at the goalkeeper surely, but beyond that it's tough to answer this question. Shots come from all over the pitch with goalkeepers and other defensive players in an infinite variety of positions. This article proposes an answer to the question of where in general to aim by asking not how to score but rather how to miss. The answer is then put to the test by examining the performance of top goal scorers in Europe and nearly every player in Major League Soccer. Missing well turns out to be a pretty good predictor of scoring on a high percentage of shots.

A Simple Model

No one can kick the ball exactly where he or she intends every time. If anyone could, their best shooting strategy would be to aim for the extreme edges of the goal, just inside the post, nearly every time. Mere mortals, however, must accept error. Decent players do not miss left more than they miss right. Good players hit the ball in the general direction they intend more often than not. In other words, the frequency distribution of shot placement is symmetric and thickest in the center. It is plausible to assume that the distribution is more or less normal.

We're now ready to illustrate my coach's "Shoot for a post" strategy on a simple one-dimensional half goal. Figure 1 shows a normal distribution centered on a goalpost. If we assume the goalkeeper can cover the goal perfectly to within one standard deviation of the shooter's distribution, then the shooter will score on 34% of shots (19% if we assume half a standard deviation). How does this technique compare with other possibilities?

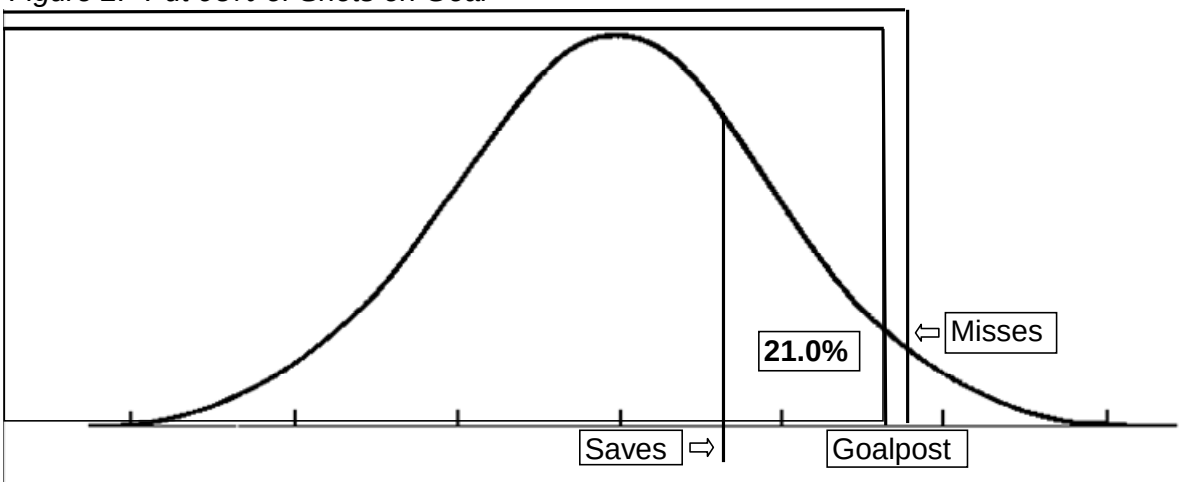
Figure 1: "Shoot for a Post"



"The most important thing of all is to get your shot on target."[1] To never miss the goal is probably an unattainable goal, but we can operationalize this theory by aiming the ball so as to miss the goal only 5% of the time. Figure 2 illustrates how this strategy fares: generating goals on only 21% of shots at the one standard deviation level (8% at half a standard deviation), again assuming the goalkeeper is perfect within his reach. To be fair, the author of this theory explains that many shots generate rebounds which in turn present scoring opportunities. His goal, quite rightly, is to maximize goals, not scoring percentages. This probably makes more sense with young players than veterans, because goalkeeping improves with age so that the number of rebounds likely declines. I'll mention this wrinkle again below, but the focus of this paper is on shot-scoring, not game-winning, percentage.

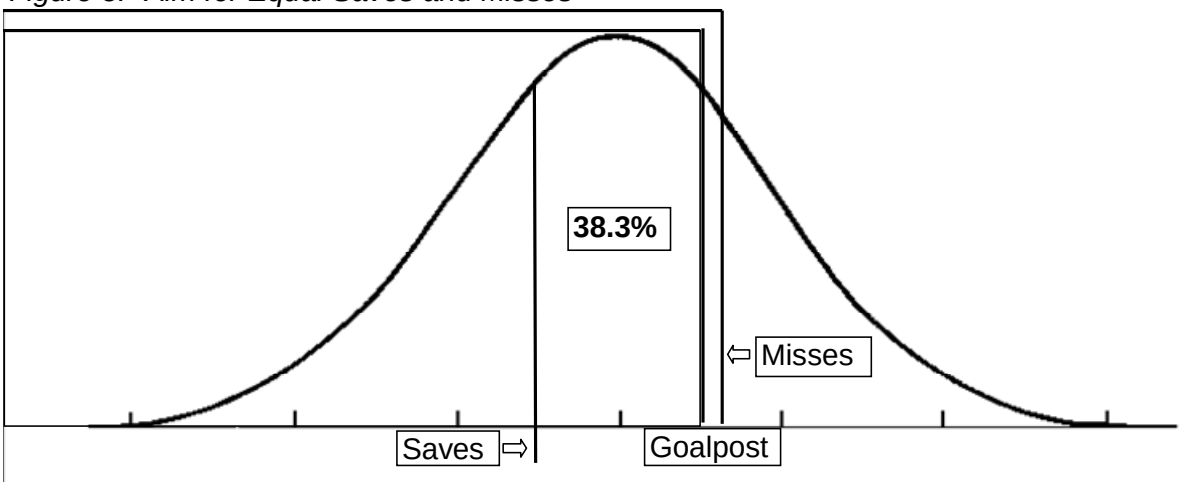
[1] Ted Buxton, Soccer Skills for Young Players (2000).

Figure 2: "Put 95% of Shots on Goal"



From these figures (if not before) it should be obvious that the optimal targeting strategy is to aim somewhere between the goalpost and the outer extent of the goalkeeper's reach. The recognition that shooting distribution is almost certainly symmetric (and assumption of goalkeeper perfection to a point) tells us where exactly to shoot: half way in between. Figure 3 moves the center of the distribution to this optimal position and generates a scoring percentage of 38% for one standard deviation (and 20% for half a standard deviation).

Figure 3: "Aim for Equal Saves and Misses"



Notice that, unlike the first two strategies, in the third there are equal shots wide of the goal and shots into the goalkeeper's arms. Thus, a simple, and easy to measure, way to describe the optimal strategy is: aim the ball so as to equalize the probability of missing the goal and permitting a goalkeeper save (i.e., a one-to-one miss-to-save ratio). It is important to note that

this result is unaffected by relaxing our assumptions regarding the precise shape of the distribution (normal) and size of scoring window (one or one-half a standard deviation). The only critical assumptions are that the distribution is symmetric and thickest in the middle. Other complicating factors, however, cannot be ignored.

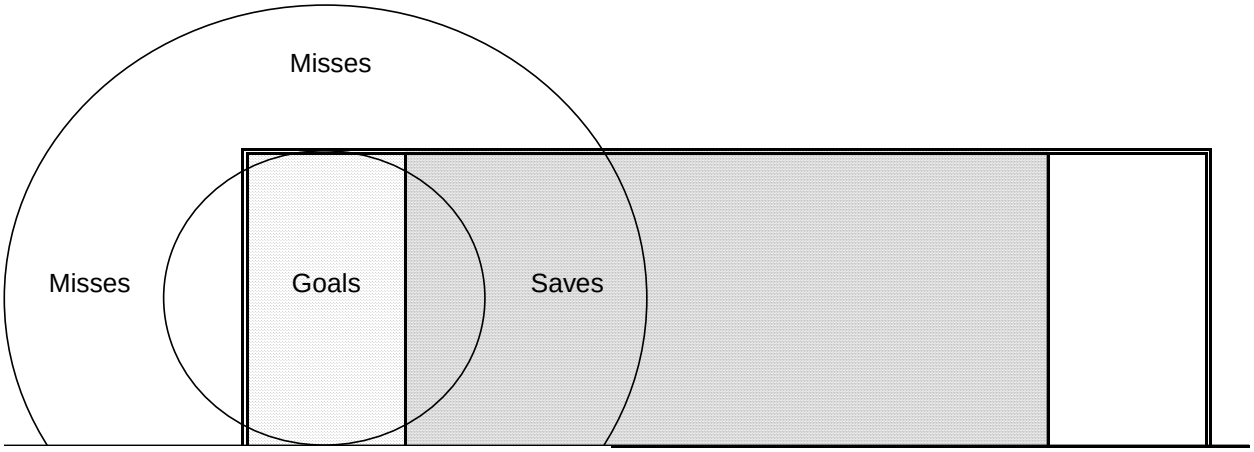
Pulling in Opposite Directions

The one-dimensional model outlined above leaves out several important factors. First, goalkeepers are not perfect walls; goalkeepers sometimes miss or bobble balls within their reach. This can result in a goal or a dangerous rebound. Focusing on goals, we can quantify the effects of porous goalkeeping. The analysis above assumed goalkeepers stop 100% of shots within their reach. Assume instead that number is 75%. How would this affect the ratio of misses to saves for the optimal scoring strategy? It would fall from one to 0.67. Graphically, one could represent this change by truncating at 75% of the normal distribution the vertical line representing the goalkeeper's reach, then adding a curve declining to the left at this constant percentage of the distribution. It may seem counterintuitive that sub-par goalkeeping leads to more saves, but one must look at the situation from the shooter's perspective. If the goalkeeper is likely to flub the play, shooters take advantage of that by steering more shots in his direction. Extra scoring space on the goalkeeper side shifts the shooter's distribution left.

Adding dimensions to the simple model reveals additional relevant factors. The second complicating factor flows from the fact that goal mouths are two dimensional. This means shooters miss high as well as wide. Adding another way to miss almost necessarily increases the miss-to-save ratio: "almost" because it is theoretically possible for players to be so accurate in their shooting that they never miss high. Figure 4 illustrates this potential (the small circle), along with the more likely case in which misses high push the miss-to-save ratio toward two (the

large circle, about 1.9 assuming a uniform distribution). An even larger circle would push the ratio higher still.

Figure 4: Effect of Misses High on Miss-to-Save Ratio



Third, because players shoot from a variety of positions on the field, two elements of a third dimension become relevant: distance from the goal and angle from dead center. Distance from the goal is closely related to missing high. The natural effect of increased distance is decreased accuracy---in other words, a higher standard deviation in shot placement. Again, this pushes the miss-to-save ratio toward misses. Angle from the center-line between shooter and the center of the goal mouth changes the shape of the target from a rectangle to what would appear to the shooter more like a trapezoid. The far post shrinks as it recedes. This is more than an optical illusion because the increased distance to the far post demands greater accuracy for scoring. The acute angle at the near post likewise gives the goalkeeper less space to cover. Thus, the important effect of angled shooting is to decrease the overall size of the target, which decreases the scoring percentage and the miss-to-save ratio. The ratio effect results from misses high remaining constant while a narrowing scoring window decreases both misses wide and saves by equal margins.

A fourth dimension and potentially complicating factor is the time it takes the ball to reach the goal. This is a function of shot speed (or power), which exerts three different effects. By reducing the distance the goalkeeper can cover, higher shot speed expands the scoring target (just the opposite effect of angled shooting). This would tend to increase the scoring percentage and the miss-to-save ratio. On the other hand, increased power means decreased accuracy and because the area of potential misses (unlike the goal) is unbounded, this pushes the miss-to-save ratio toward misses. The third effect of shot speed is to make the goalkeeper more porous. It is more difficult to catch a ball coming at high speed. As argued above, the impact of goalkeeper error is to increase saves relative to misses. Thus, the net effect of increased shot speed on the miss-to-save ratio is ambiguous: the bigger target and decreased accuracy pushes toward misses; powering through the goalkeeper pushes toward saves.

A final complication is goalkeeper behavior. The economics literature has examined the interaction between shooter and goalkeeper in some depth in the context of penalty kicks.[2] In that setting, a critical element of shooting strategy is tricking the goalkeeper into diving the wrong way. This game theory component of shooting is likely much less important in the ordinary run of play, where most shots are scored too quickly for player reflection or faints. Still, the clear effect of deception is to put more shots on goal. As for the miss-to-save ratio, the impact is precisely the same as the porous goalkeeper effect above. Indeed, fooling the goalkeeper is merely one way to score goals that were within the goalkeeper's grasp.

What predictions then can we make about the aiming strategies of professional soccer players? The simple model shown in Figure 3 suggests that players who score on a high percentage of shots will tend to miss about as frequently as their shots are saved. In other words,

[2] P.-A. Chiappori, S. Levitt, & T. Groseclose, "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer," American Economic Review 92(4), 1138 (2002).

scoring percentage will decrease as the ratio of misses to saves deviates from one. If shot placement indeed follows a normal distribution, we would expect the scoring percentage drop-off to be gradual near one, rapid in middle values, and slow as the deviation from one is greater.

But what about the various real world effects described above? There are three possibilities: (1) they cancel one another perfectly and the optimal miss-to-save ratio remains one; (2) the porous goalkeeper effects dominate and saves outnumber misses at peak scoring percentage; or (3) misses high push the optimal miss-to-save ratio above one. I think the third option is the most likely, but I do not think the best miss-to-save ratio will be a great deal higher than one. The geometry of Figure 4 is compelling, but high percentage shots are likely to be from close range, where misses high are unlikely. And my impression is that professional goalkeepers usually make saves when given the opportunity. Deception is almost certainly not a dominant shooting strategy in the run of play. In other words, I predict that the simple model will hold sway, with a modest adjustment for shots over the cross-bar generating a miss-to-save ratio slightly above one.

Results

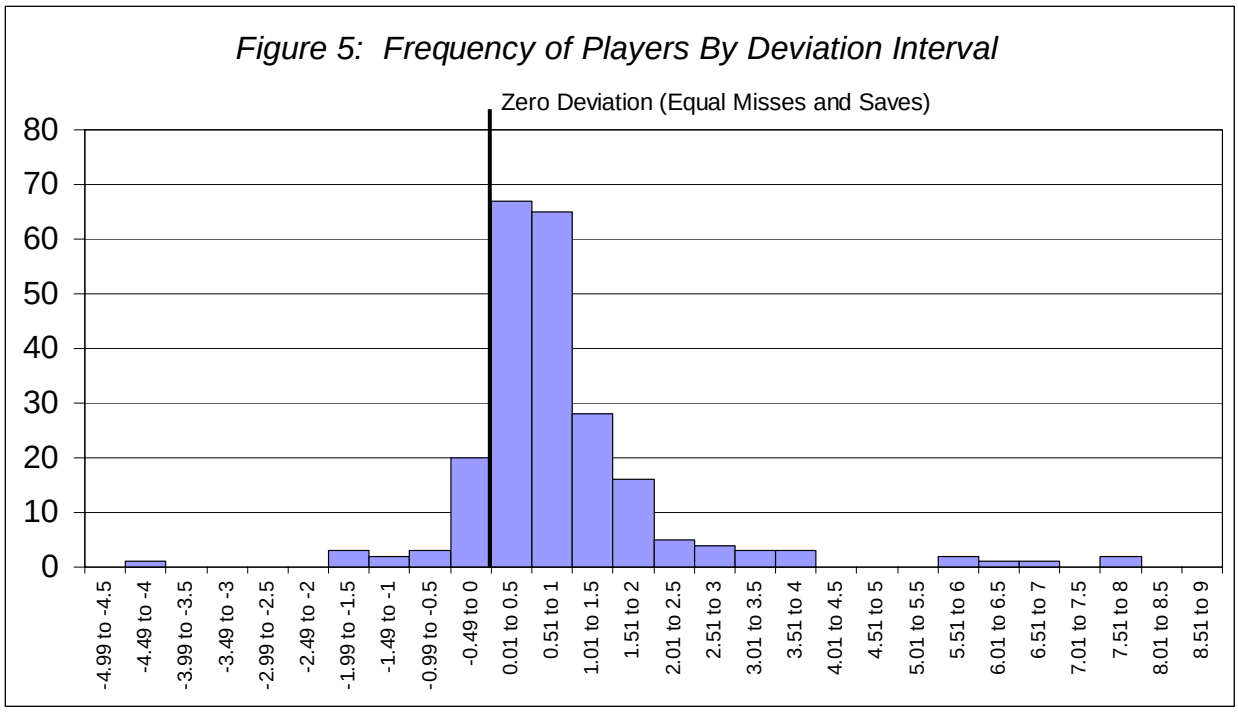
To test these hypotheses, I started with the best players in the world. Europe is home to the best soccer teams. In 2008, Forbes Magazine reported that all ten of the top ten teams in value were European.[3] The Union of European Football Associations ("UEFA") Champions League is Europe's premier club competition. In the 2007-08 season, 39 players in the league scored three or more goals. Twenty of these scored on 25% or more of their shots. Among these 20, there were precisely the same number of saves and misses wide: 91 each. This suggests that the most efficient scorers in the world apply the strategy depicted in Figure 3.

[3] "The Most Valuable Soccer Teams," http://www.forbes.com/2008/04/30/valuable-soccer-teams-biz-soccer08-cx_jg_pm_0430soccer_land.html (visited Aug. 9, 2008).

But looking at one extreme tail of the distribution of scorers may miss broader trends and differences among individual players. The UEFA data also reflect few shots per player. To correct these problems I gathered career data on each player active in the United States's top professional league, Major League Soccer ("MLS"), as of the all-star break of the 2008 season. I omitted goalkeepers and players who had zero misses or saves. (It was impossible to calculate a miss-to-save ratio for such players.) The resulting dataset consisted of 226 players out of 381 in the league. It is this MLS dataset I'll be analyzing going forward.

In order to put values below one on the same scale as values above one, I converted the miss-to-save ratios to deviations from a ratio of one by inverting ratios less than one, making negative those values, then adding one. The conversion for miss-to-save ratios greater than one was simply to subtract one. Thus, the new deviation variable took on a value of zero for players with equal misses and saves, a positive value for players with more misses, and a negative value for players with more saves. A player with twice as many misses as saves, for example, had a deviation value of positive one; twice as many saves as misses generated a negative one value.

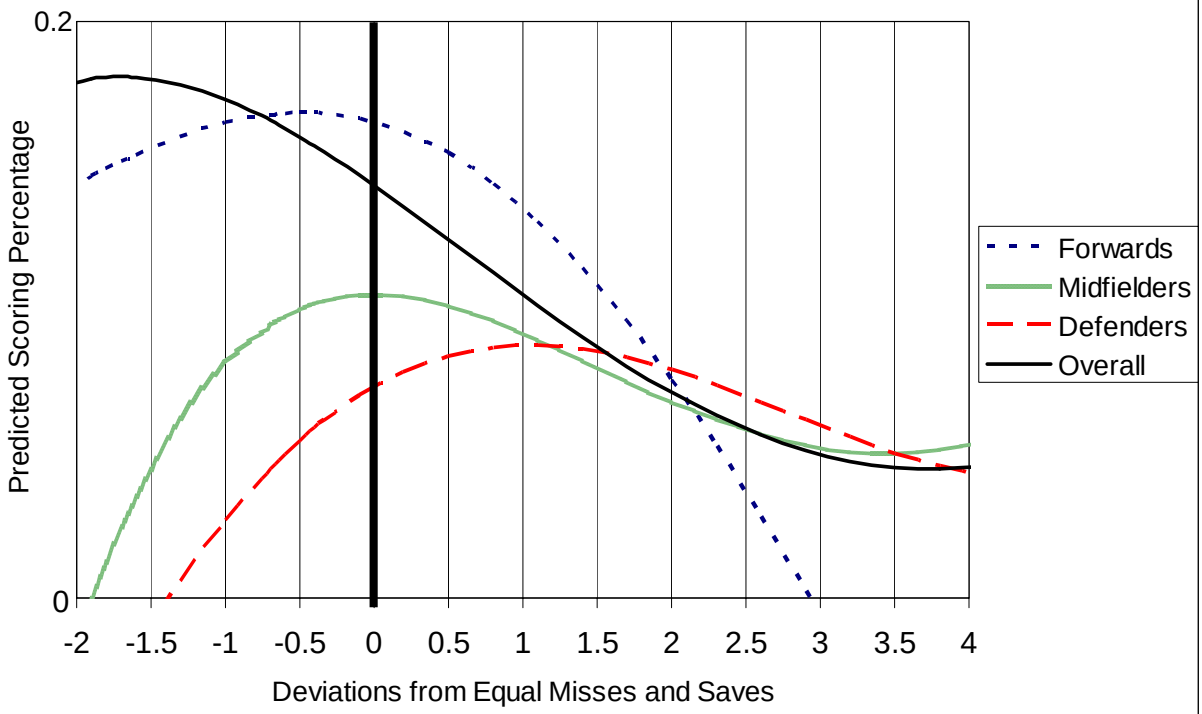
The frequency distribution of the deviation variable is itself quite telling. As you can see in Figure 5, the overwhelming majority of players had deviations from equal misses and saves of between -0.5 and +2. Clearly, there is a strong central tendency toward equality. There is, however, a noticeable shift to the right, suggesting slightly more misses than saves. That professional players shoot in this way is reassuring for my predictions, but this histogram tells us little about whether a one-to-one miss-to-save ratio is the optimal shooting strategy within the league. To answer that question, we turn to scoring percentage data.



The average scoring percentage for all players in the dataset was 11.8%, weighted by shots taken. Figure 6 shows weighted regression lines for scoring percentage against deviations from a one-to-one miss-to-save ratio. I included on the right-hand side of the equation deviation along with squared, cubed, and fourth-power terms. The higher order terms are to allow for three minima and maxima. My prediction again was that the line would have minima near the bottom and top ends of the deviation range and a maximum just above 0.

The overall data, represented on the graph by the solid black line, do not bear out this prediction. The observed local maximum was about -1.7, which corresponds to a ratio of 2.7 saves for each miss. Notably, the deviation coefficient was negative and highly significant ($p<0.001$). The cubed term was positive and marginally significant ($p=0.092$), but the squared and fourth-power terms were not statistically significant.

Figure 6: Relationship of Deviations to Scoring Percentage



What explains the preponderance of saves among high percentage goal scorers? It would appear that I underestimated the fallibility of goalkeepers and the accuracy (and perhaps power) of professional goal scorers. As explained above, a porous goalkeeper will push the miss-to-save ratio toward saves because shooters will place more balls on the goal mouth. On the other hand, high shooter accuracy will tend to push down misses faster than saves since there are two ways to miss (high and wide).

Both of these effects will be most pronounced for shots from close range: goalkeepers have less time to react and shooters become much less likely to miss, either high or wide. One way to see if close range goals are driving the observed result is to divide the dataset by player position. Forwards are more likely to shoot from close range, defenders from distance, and midfielders from somewhere in between. My new prediction is that the optimal deviation will be lowest for forwards and highest for defenders. The data do seem to support this hypothesis.

Figure 6 shows that predicted scoring percentage peaks at a deviation of around -0.5 for forwards (1.5 saves per miss), 0 for midfielders (equal misses and saves), and 1 for defenders (2 misses per save). It should be emphasized, however, that the deviation and higher order coefficients were not significant in any of these three position-specific regressions, so these results are suggestive at best.

Still, these findings, when combined with a few autobiographical facts, may explain why my predictions were off the mark. I have played soccer since I was very young, always as a defender. I have scored precisely twice in competitive play, both times from long distance. The geometry of missing high and wide naturally looms large in my psyche. I clearly don't know how to score from the forward position. For the sub-sets of midfielders and defenders my hypotheses were much nearer to the target. My coach's advice, by the way, may have been well directed after all. The scoring window at the long distance I was likely to shoot from was going to be narrow sliver by the goalpost. The optimal strategy may have been to aim for the center of that space, but the post, a clearly defined target, may actually have been a more efficient proxy. Better players in better scoring positions, however, would do well to ignore my coach's advice.

Postscript: A Notable Outlier

Some players do not fit into my theory. A handful of players have miss-to-save ratios at or above four and also have above average scoring percentages. All but one of these data points reflect fewer than 20 shots. The only player with a miss-to-save ratio above four (8.67) and more than 20 shots (34) is David Beckham, and he had an above-average scoring percentage of 14.7%. In Major League Soccer, Beckham is exceptional in many ways. "He has twice been runner-up for FIFA World Player of the Year, and in 2004 was the world's highest-paid

footballer."[4] His 2008 guaranteed compensation was $6.5 million, 50 times the MLS average and more than double the next highest paid player.[5]

What can Beckham's numbers tell us? One implication could be that the other, inferior MLS players follow a sub-optimal scoring strategy. The best players in the world, like Beckham, may score in a systematically different way. The initial examination of UEFA data---with equal saves and misses at the highest level of soccer---suggests that this is not the case. Looking more closely at Beckham's pre-MLS career also implies that the later part of Beckham's career is exceptional. Before joining MLS, four seasons of shooting statistics are readily available.[6] In the 2002/03 and 2004/05 seasons with Manchester United and Real Madrid, respectively, Beckham had slightly more saves than misses (deviations of -1.28 and -1.25) along with scoring percentages of 9.5% and 12.9%. The following two seasons with Real Madrid were much different: Beckham's miss-to-save ratio soared (deviations of 14.7 and 16.5) and his scoring percentage declined (6.0% and 7.9%). In other words, over four seasons in Europe, Beckham's shooting efficiency was highest when his shots generated nearly equal misses and saves.

What might explain the change in Beckham's shooting strategy? A loss of running speed is one possibility. A player who can move quicker than defenders is more likely to find himself with a scoring opportunity close to the goal. Why take a shot from distance if you can zoom around another defender and have a closer chance? Similarly, fast players can get to the end of crosses, which generate high probability scoring chances. Both effects would push the miss-to-save ratio toward one since faster players shooting from closer and finishing crosses would likely

[4] http://en.wikipedia.org/wiki/David_Beckham (visited Aug. 10, 2008).
[5] http://nbcsports.msnbc.com/id/23961529/adidas (visited Aug. 10, 2008).
[6] http://soccernet.espn.go.com/players/stats?id=8880&cc=5901 (visited Aug. 10, 2008).

put fewer shots over the cross-bar than slower players. The question becomes: why didn't Beckham's scoring percentage fall more dramatically with age?

One possible answer lies in dead-ball scenarios. Beckham is famous for his free-kicks, for bending balls around walls of opponents into the extreme corners of the goal mouth.[7] Saves on free-kicks are relatively rare. The wall replaces the goalkeeper as the primary obstacle between the ball and the goal-mouth. Shots into the wall will be coded as misses rather than saves. Beckham's declining speed almost surely means that a higher percentage of his shots are coming on free-kicks, which would naturally bring down his shots saved. Because he is so effective on free-kicks, this shift toward misses rather than saves may not have been accompanied by the decrease in scoring percentage one might otherwise predict.

Moving to MLS likely increased Beckham's scoring percentage, reflecting inferior competition. But surely more is at issue. Beckham strikes the ball harder and more accurately than perhaps any other player in the world. Both characteristics tend to increase scoring percentage. The net effect on miss-to-save ratio is unclear: hitting the ball harder would tend to reduce accuracy and thus increase the miss-to-save ratio, whereas greater accuracy pushes the ratio toward one. Beckham's abilities almost certainly change his shot selection. He has a better chance of scoring on long distance shots than other players, so he likely lets rip from distance, even more so as he has aged and slowed. This unequivocally increases his miss-to-save ratio, since longer shots, even by a relatively accurate shooter, are more likely to go over the cross-bar.

Another explanation for Beckham's high miss-to-save ratio in MLS is that, especially as his career advances, he may value spectacular goals more than ordinary goals, on free-kicks and more generally. The spectacular goal glances off the post and in at rocket speed. Beckham may employ a shooting strategy more consistent with my youth soccer coach's "shoot for a post."

[7] Bend It Like Beckham (2002).

What is perhaps remarkable is that Beckham is so accurate a shooter that he can get away with it.[8]

[8] "Former manager Sir Alex Ferguson said that he 'practised with a discipline to achieve an accuracy that other players wouldn't care about.'" http://en.wikipedia.org/wiki/David_Beckham (visited Aug. 10, 2008).