

## **World Cup 2006:** **Predicting a Little Box of Surprises**

So we've come full circle. Four years ago Dectech began modelling football for the World Cup 2002. Since then, we have calculated odds for nearly 10,000 domestic and international matches, researched a range of different model improvements and run a football column in the Times<sup>a</sup>.

As this brief explains, our forecasts outperform both newspaper commentators and bookmakers' odds. So who do we think is going to win this summer? Well, with still a few friendlies to go, Brazil is the current favourite with a 13% chance. But that still leaves an 87% chance for someone else. Maybe even Trinidad...

Dectech doesn't really have an office sweepstake for the World Cup. What's the point? It's like asking Gordon Ramsay where he's having the staff Christmas party or Donatella Versace what she's wearing to the Oscars. However, don't let this discourage you from using these predictions in your own office competition. In fact, this could be your best shot at achieving that elusive victory over the Fixed Income Desk / IT Department / Non-Executive Directors (delete as appropriate).

Brazilians are prone to pointing out that "football is a little box of surprises", and you only have to look back at the Greek victory in Euro 2004 to see what they mean. Indeed, football is certainly one of the hardest prediction problems we've ever tackled. Nevertheless, even if we still fall short of outright clairvoyance there is evidence that the probabilities provided in this brief are the most accurate published. As we keep pointing out, it's quite hard to forecast the world, but it's quite easy to beat people's intuitive judgements about it.

In this brief we first introduce the model, its current assessment of the teams and its historical performance. Then we detail our Cup forecasts. Finally, we consider England's potential journey to the final, with or without Wayne Rooney<sup>b</sup>. Meanwhile, for the latest forecasts as the tournament evolves, details on any specific teams or games, and comparisons to bookmakers' odds, please stop by our website, [www.dectech.org](http://www.dectech.org), and click on "football" in the menu bar.

**Das Modell**

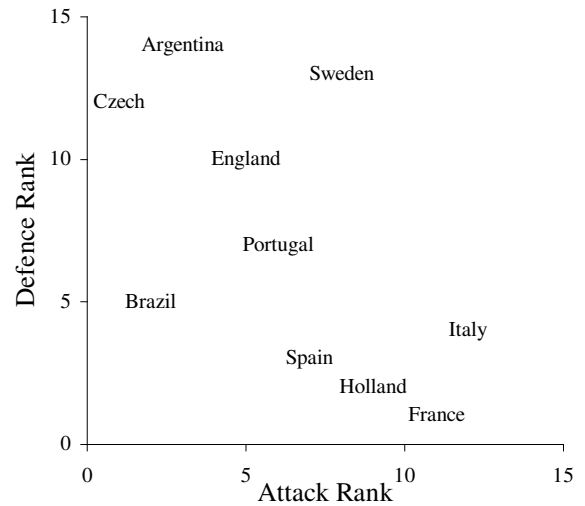
The model makes forecasts based on past team performance. Essentially, we take about 4,500 international games, including World Cup qualifiers and friendlies, and see how the different teams fared against each other, with more weight given to recent matches.

Intuitively, you can infer the likely outcome of a match between Team A and Team B by carefully blending what happened last time they met with how the two teams have performed more recently against other teams they've played in common (or they've played someone who has, etc.). This is basically what the model does<sup>c</sup>. The outputs have three main applications that are discussed over the remainder of the brief – team ranking, match forecasting and tournament simulation.

First, the immediate outputs of the model are estimates of the pitch action strengths of different teams (e.g. how many goals they can expect to score against an average opponent). These outputs can be used to rank the attacking and defensive capabilities of each team and these are reported in Figure 1 for the ten strongest sides in this year's Cup. The closer you

get to the bottom left hand corner of the figure, the better. For example, Brazil is currently one of the strongest all round team since they rank second in the world for attack and fifth for defence. The world's best attacking team, the Czech Republic, is more lopsided since they rank 12<sup>th</sup> in defence. As such, they don't rate so highly overall.

Figure 1. Attack and Defence Ranks



England's attack puts it in the Cup top ten, but the side is comparatively weak at the back. Out of interest, the three strongest teams that didn't qualify for this year's tournament are (best first) Ireland, Denmark and Romania.

**Match Predictions**

Next we use these team pitch characteristics to forecast specific matches. Because there are estimates for every team, forecasts can be made for any match. Figure 2 shows the current predictions for England's game with Sweden on 20<sup>th</sup> June.

Figure 2. England Vs Sweden Predictions

Percent Chance	England				
	0	1	2	3	4
0	6.0	8.9	6.5	3.2	1.2
1	8.1	11.9	8.7	4.3	1.6
Sweden 2	5.4	7.9	5.8	2.9	1.1
3	2.4	3.5	2.6	1.3	0.5
4	0.8	1.2	0.9	0.4	0.2
<b>Sweden Win</b>		<b>Draw</b>		<b>England Win</b>	
34.5%		25.2%		40.4%	

The figure shows that the most likely outcome is a 1-1 draw, with a probability of 12%. However, adding up the relevantly coloured full time scores shows that the most likely overall result is an England win at 40% or

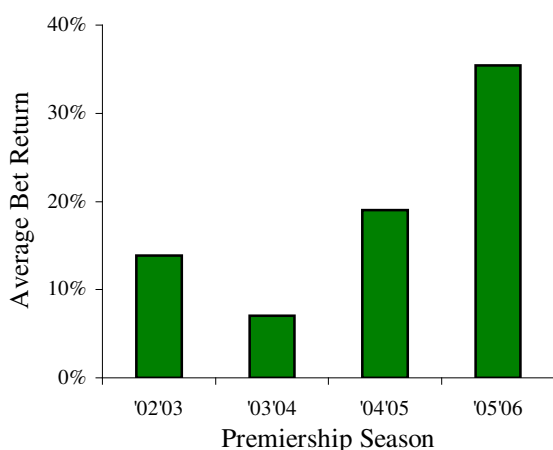
odds of 3/2. Likewise, the full time score probability distribution can be used to derive other odds such as Asian handicaps, winning margins, and total goals.

### Predictive Accuracy

As mentioned earlier, there is good evidence that these predictions are more accurate than any of the others in circulation. For instance, backtests against results in prior domestic seasons and internationals demonstrate remarkable accuracy. Yet the acid test is how well the published predictions have performed.

One comparison is against other newspaper tipsters. Generally, these tipsters gauge the correct match outcome around 43% of the time<sup>d</sup>. By contrast, over the last four English Premiership seasons our predictions correctly called 53% of matches. A second test is shown in Figure 3. When you compare our predictions to bookmakers' odds you typically find that we disagree on about a quarter of the games. The graph shows the average return obtained from betting against the bookmakers on these mis-priced matches.

Figure 3. Four Seasons of Premiership Forecasts



Hence, for the 102 matches where our forecasts diverged from bookmakers' odds in '04'05, placing £1.00 on one of the matches yielded average winnings of £1.20. Clearly, by compounding matches over the season and choosing best odds, the true annual returns of these forecasts can get much higher (for example, in '04'05 we actually made a return of about 75%). Equivalent analysis for other leagues and international tournaments yields a similar picture.

So how is this possible? Our previous brief on player valuation described how statistics can be used in most domains to identify anomalies in human judgement. People who try to forecast football with brains alone simply cannot process all the information contained in thousands of games between hundreds of teams. To do that, you need algebra and a laptop.

### Never Mind That, Who's Going to Win?

The final application of the model is to take individual match predictions based on team rankings and then aggregate them into forecasts for the overall tournament. This is typically done using techniques like Monte Carlo simulation. Results from thousands of such tournament simulations for this year's competition appear below.

Figure 4. World Cup 2006

Country	2 <sup>nd</sup> Round (%)	Win (%)
Brazil	85.0	13.1
France	89.5	11.1
Germany	90.8	10.9
Netherlands	77.4	10.7
Spain	90.3	9.9
Czech Rep	71.9	6.9
Portugal	85.5	6.2
Italy	67.1	5.4
England	79.7	5.0
Argentina	62.1	4.4
Sweden	73.9	3.6
USA	51.9	2.3
Croatia	53.3	1.9
Mexico	64.7	1.7
Poland	56.5	1.2
Switzerland	52.5	0.7
Serbia	29.4	0.7
Ivory Coast	31.1	0.6
Ukraine	45.8	0.6
Paraguay	42.0	0.5
Australia	32.1	0.5
Tunisia	51.7	0.5
South Korea	40.3	0.4
Japan	29.6	0.3
Ecuador	34.0	0.3
Iran	30.3	0.1
Costa Rica	18.7	0.0
Angola	19.6	0.0
Ghana	9.1	0.0
Togo	17.7	0.0
Saudi Arabia	12.2	0.0
Trinidad	4.4	0.0

Hence, for example, the chart shows how Switzerland's chance of emerging from Group G is 53% though their chances of getting to the final and winning are slightly under 1%. Overall, Brazil is the current favourite with a 13% chance of winning. However that still leaves an 87% chance for someone else to lift the Cup. Indeed, even those teams that have a "0.0%" chance actually have some chance, albeit very small<sup>e</sup>. Summing the bottom of the table, there is a 10% chance that someone who currently has less than a 2% chance of winning will emerge the victor - "the little box of surprises" effect.

In identifying Brazil as the favourites, the model is in qualitative agreement with bookmakers. But comparing Figure 4 to current odds does highlight substantial differences. For instance, Brazil's current odds to win are 4/1 implying that they have a 20% chance, which seems optimistic for any knock-out tournament. Likewise, the patriotic betting of proud nationals has placed England at 8/1 or 11%, rather ahead of the more sober analysis presented here. Conversely, the Czechs, French and Dutch are all significantly undervalued.

**Path To The Prize**

It's been 40 years since England got their hands on the trophy. What challenges must they overcome to repeat such a triumph? One consequence of running the tournament simulations is that we can also track the most likely opponents any team would face, given they reach a certain stage of the competition. These estimates are shown in Figure 5 for England. For all the World Cup teams each column sums to 100%, but the table only shows those that England currently have a greater than 10% chance of playing.

Figure 5. England's Likely Opponents

Country	Last Sixteen	Quarter Final	Semi Final	Final
Germany	40.3	-	-	13.1
Poland	30.3	-	-	1.4
Ecuador	19.0	-	-	0.4
Costa Rica	10.4	-	-	0.1
Portugal	-	25.7	-	5.7
Holland	-	23.7	-	10.7
Argentina	-	18.0	-	4.5
Mexico	-	12.7	-	2.2
Brazil	-	-	19.8	9.6
Spain	-	-	19.2	7.6
France	-	-	13.8	11.7
Czech Rep	-	-	10.8	7.1

The figure provides some insights into the workings of the competition and the obstacles England might face. The second round starts with the graduates from Group B playing Group A. Hence, the first column combines England's chances of finishing first or second in their group with the various chances of different Group A outcomes. The upshot is that if England get through, their most likely opponent is the old nemesis Germany. Now, with total objectivity, we currently rate England as better than Germany. However, home advantage is large for internationals, adding an extra 0.58 goals on average for the locals. So on that basis we estimate England's actual chances of winning such an encounter to be 41%.

The next three columns of Figure 5 then detail the various other opponents England will possibly have to beat if they are to win the competition. The quarter-

finals would pit them against a team that emerges from Groups C or D, most likely Portugal or Holland. This would then be followed by a semi-final with Brazil or Spain. Again, details of the likelihoods of different future opponents are provided for every team on our website and will be updated during the competition as the actual results come in.

**Conclusion**

As Bill Shankly famously remarked<sup>f</sup>, "Some people believe football is a matter of life and death. I'm very disappointed with that attitude. I can assure you it is much, much more important than that." In a world that often seems divided, we need to remember that there are also subjects, like football, that unite us. No matter how small, in the World Cup everyone is in with a chance, and we'll drink to that.

**References and Footnotes**

- a. The Fink Tank column first appeared in the UK's Times newspaper on 15<sup>th</sup> November, 2002 and has run each Saturday during the English League season ever since. The column applies quantitative analysis to topical issues in football and gives predictions for the weekend games.
- b. Based on current performance, Table 4 shows that England's chances of winning the Cup are 5.0%. However, if you substitute the average England team from 2002, before Wayne Rooney was capped, that chance falls to 4.9%. We are currently doing more research on this, but we have previously noted that whilst Rooney was a gifted player at Everton, the *team* actually played better without him.
- c. For the technically minded, the forecasts are built using maximum likelihood estimation based on a model that assigns parameters to teams together with other parameters for items such as home advantage. For an introduction to this kind of modelling in football see Dixon, M. J. & Coles, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46, pp: 265-280.
- d. Forrest, D. & Simmons, R. (2000) Forecasting sport: The behaviour and performance of football tipsters. *International Journal of Forecasting* 16, pp: 317-331. They measured the proportion of matches called successfully as Daily Mail 43%, Mirror 41%, and Times 43% (before the Fink Tank, naturally).
- e. The more simulations we carry out, the more we can account for extremely unlikely events. For example, Trinidad & Tobago's chance of winning the World Cup is actually 0.0009% - a figure only revealed after hundreds of thousands of simulations.
- f. Bill Shankly was a Scotland international player and manager of Liverpool. Interestingly, he has been proved right since heart attacks increase in males when their football team loses at home. See Kirkup, W & Merrick, D. W. (2003) A matter of life and death: Population mortality and football results *Journal of Epidemiology and Community Health* 57, pp: 429-432.