Penalty shootouts: are estimates on the spot?

James Jackson

October 21, 2022

On Sunday 20th November the 2022 FIFA World Cup will get underway in Qatar. While humanitarian issues have dominated the build-up thus far, attention will soon turn to the football, and, inevitably, to penalty shootouts.

England, in particular, have an infamously poor track-record with penalty shootouts. Their most recent shootout defeat came against Italy in the final of the 2020 European Championships. Afterwards, England manager Gareth Southgate faced criticism because, according to pre- and post-match interviews, penalty takers were selected based on performance in practice, rather than factors such as players' experience, positions and the minutes they played in the final.

This article explores the extent to which such factors influence penalty taking, by revisiting a paper by Jordet et al. (2007). Their analyses have then been repeated using more recent data; the results of which have been used to take a closer look at the England-Italy penalty shootout.

What is a penalty shoot out?

Penalty shootouts are typically used to decide the winner in knockout rounds of football tournaments, whenever the match ends in a draw. Quite simply, whichever team scores the most penalties wins. A penalty itself is a sole attempt at goal from 12 yards. A penalty shootout initially comprises five penalties per team, where each kick is taken by a different player. If both teams score the same number, the shootout enters a "sudden-death" phase, where it progresses until one team misses and the other scores.

The Jordet et al. (2007) study

Jordet et al. (2007) analysed data relating to all kicks taken in penalty shootouts in three of the leading international football tournaments - the World Cup, the European Championships and the Copa America - between 1976 and 2004. They computed odds ratios to compare how the odds of scoring changes among different categories for variables such as players' positions and ages; these variables are listed in Table 1.

Table 1: The five variables considered in the original Jordet et al. study.

Variable	Brief description
Tournament	Copa America, European Championships, World Cup
Kick number	1, 2, 3, 4, 5 and 6[1]
Position	Forward, Midfielder, Defender
Playing time (minutes)	1-30, 31-90, 91-120
Age	18-22, 22-28, 29+ (years)

[1] This includes all "sudden-death" penalties.

The *odds* of scoring a penalty is just the probability of scoring divided by the probability of missing. The *odds ratio (OR)* compares the odds of scoring a penalty in a given category, relative to a reference category. For example, the OR for those aged 18-22 gives the odds relative to the odds for those aged 29 or above (the reference category). *Marginal* ORs were computed, that is, each variable was considered separately. An alternative would have been to calculate *conditional* ORs. Both types of OR can be obtained via logistic regression; see box for the distinction between the two. In this instance, the conditional ORs were similar to the marginal ORs, though they have the potential to be very different. This phenomenon, known as Simpson's paradox, is named after Edward H. Simpson, who, among other things, was a codebreaker at Bletchley Park during the Second World War.

Marginal vs. Conditional ORs

The *marginal* ORs can be obtained by fitting separate logistic regression models (one for each variable - age (X_1), position (X_2), minutes played (X_3)) where the outcome of the penalty Y (scored or missed) is the response and the variable in question is the sole predictor. For example, for age:

$$Y \sim \text{Bernoulli}(p)$$
$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1.$$

The marginal log-ORs for age are given as the estimates of the regression coefficient β_1.

The *conditional* ORs can be obtained by fitting a single logistic regression model where again the response is the outcome of the penalty, but where age, position and minutes played are all predictor variables.

$$Y \sim \text{Bernoulli}(p)$$
$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3.$$

The conditional log-ORs for age are again given as the estimate of β_1.

Several fascinating point estimates - that is, the best supported value based on the data - ensued; these are presented on the left-hand side of Table 2. The odds that players taking the first penalty in the shootout were successful were 3.58 times greater than those kicking in the sudden-death phase of the shootout; the odds that younger players (aged 18-22) scored were 1.62 times greater than older players (aged 29+); and the odds that extra-time substitutes (those whose playing time was 1-30 minutes) scored were 1.84 times greater than those who started the match. The study also found that forwards were more likely to score than defenders, and penalties were less likely to be scored in the World Cup than in the other tournaments.

One can propose reasons why such relationships exist. For example, more able penalty takers tend to kick earlier in the shootout, whereas those less skilled are called upon to take the sudden-death penalties. Moreover, substitutes' success could be attributed to reduced fatigue; or they may have been substituted on *because* they are good penalty takers. Finally, younger players' success could be attributed to lower stress levels: they are neither scared of failure nor scarred by previous failures.

However, the ORs' standard errors were large, hence confidence intervals wide. This was because of the relatively small sample size of the data used - simply not enough penalties had been taken. Thus many of the point estimates were found not to be statistically significant, that is, there was insufficient evidence to suggest the ORs were not equal to 1. This means it was not possible to conclude that there were any differences between the categories considered.

As a side note, Jordet et al. made an implicit assumption: that the data constituted a simple random sample drawn from a super-population, that is, standard errors were not adjusted to account for a finite population. This super-population can be imagined as encompassing all future penalties in international football tournaments not yet observed.

Repeating the study with more recent data

Since the original Jordet et al. study there have, of course, been many more penalty shootouts. To facilitate a comparison with the results obtained by Jordet et al. - and identify recent trends - the analysis has been repeated using post-2004 data, in addition to data from the Africa Cup of Nations, another leading international football tournament. The data were obtained through the extensive database available on Wikipedia.

The results from the post-2004 data have been given alongside the pre-2004 data in Table 2, and also presented as bar charts in Figure 1. Several point estimates changed substantially; for example, for the *position* variable, the OR for forwards fell from 1.76 to 1.01, while the OR for midfielders fell to 0.83, below 1. On the other hand, the ORs for the *age* variable were remarkably similar, showing that younger players are more likely to score. Yet, although the sample size is larger and standard errors generally smaller, it is still not sufficient for any of the estimates to become significant.

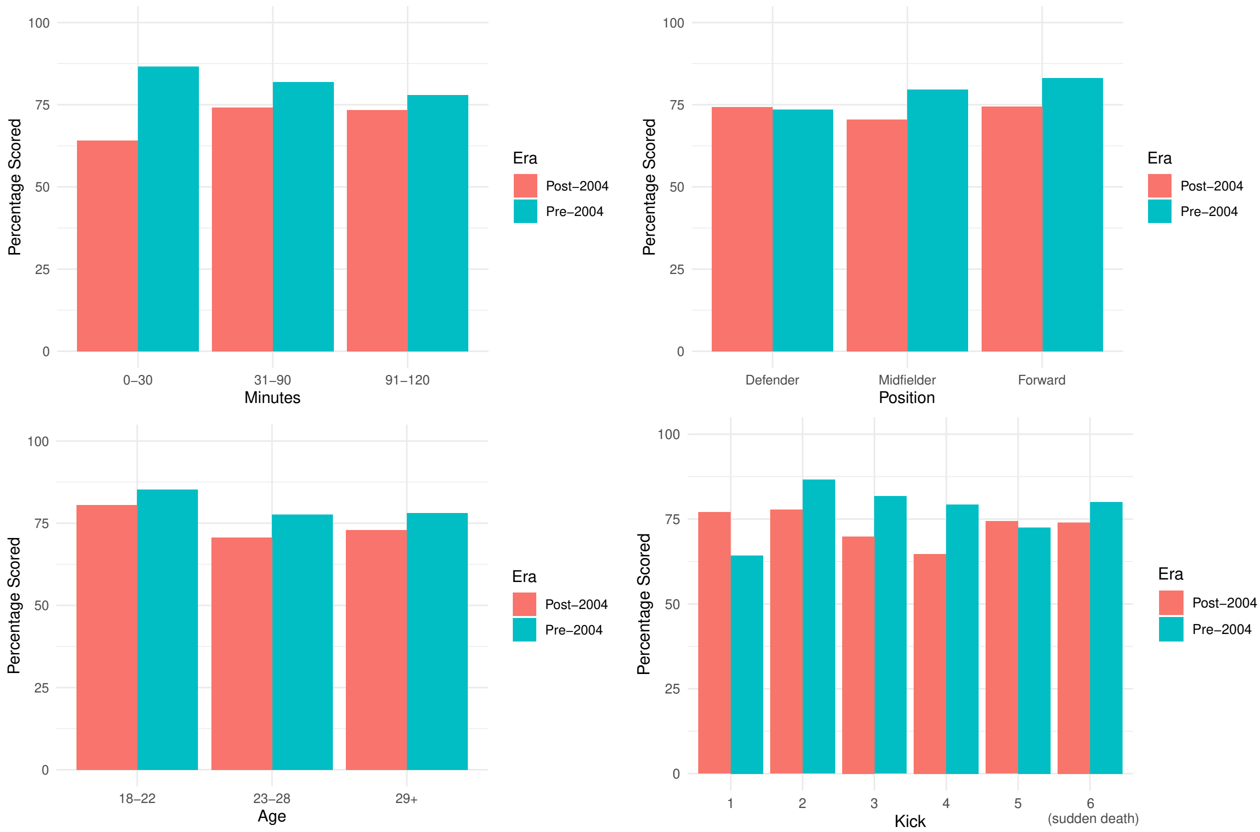

Figure 1: The percentage of spot kicks scored. The blue bars give the percentages from the original Jordet et al. study (pre-2004 tournaments). The red bars provide the percentages from post-2004 tournaments.

Table 2: Sample sizes, odds ratios, confidence intervals and p-values for the pre-2004 data and post-2004 data.

	pre-2004 data $n = 409$			**post-2004 data** $n = 648$		
Variable	OR	C.I.	p	OR	C.I.	p
Tournament						
Africa Cup of Nations	-	-	-	1.57	(0.98, 2.53)	0.06
Copa America	1.93	(1.09, 3.41)	0.024	1.08	(0.65, 1.82)	0.76
European Championship	2.21	(1.21, 4.03)	0.010	0.95	(0.55, 1.65)	0.85
World Cup	1	Ref.	Ref.	1	Ref.	Ref.
Kick number						
Kick #1	3.58	(1.32, 9.75)	0.012	1.18	(0.60, 2.29)	0.63
Kick #2	2.48	(0.96, 6.45)	0.062	1.23	(0.63, 2.41)	0.54
Kick #3	2.12	(0.83, 5.44)	0.116	0.81	(0.43, 1.56)	0.53
Kick #4	1.46	(0.59, 3.66)	0.414	0.65	(0.34, 1.23)	0.18
Kick #5	2.22	(0.80, 6.14)	0.124	1.02	(0.49, 2.12)	0.96
Kick #6 *	1	Ref.	Ref.	1	Ref.	Ref.
Positional role						
Forward	1.76	(0.95, 3.28)	0.073	1.01	(0.64,1.58)	0.98
Midfielder	1.41	(0.81, 2.45)	0.230	0.83	(0.54, 1.27)	0.38
Goalkeeper/Defender	1	Ref.	Ref.	1	Ref.	Ref.
Playing time						
1-30 mins	1.84	(0.41, 8.34)	0.430	1.54	(0.85, 2.82)	0.16
31-90 mins	1.28	(0.67, 2.47)	0.455	1.60	(0.83, 3.09)	0.16
91-120 mins	1	Ref.	Ref.	1	Ref.	Ref.
Age						
18-22 years	1.62	(0.70, 3.77)	0.263	1.54	(0.85, 2.78)	0.15
23-28 years	0.97	(0.56, 1.69)	0.978	0.89	(0.60, 1.32)	0.56
29+ years	1	Ref.	Ref.	1	Ref.	Ref.

* This includes all “sudden-death” penalties.

A closer look at the England vs. Italy shootout

Going back to England and Italy’s shootout in the final of the Euros, how do the two teams’ choice of penalty takers compare, based on the data? And given the teams’ choice of penalty takers, who should have won?

A penalty shooutout can be modelled mathematically very simply. It can be viewed as a series of independent Bernoulli random variables. A Bernoulli random variable - the classic example being tossing a coin - has a probability of success, say p. In this instance, of course, p is the probability that a player scores their penalty, and this can be estimated from their age, playing time and position.

Specifically, logistic regression was used to estimate “p” for the players selected by England (Kane, Maguire, Rashford, Sancho and Saka) and Italy (Berardi, Belotti, Bonucci, Bernardeschi and Jorginho), assuming the variables to be independent. Given the two teams’ players’ scoring probabilities, Monte Carlo simulation was then used to establish the probability that England or Italy would go on to win the shootout. Essentially, Monte Carlo simulation utilises the frequentist view of probability: the proportion of times an outcome occurs in the long-run. That is, the penalty shootout was re-run, virtually, a million times using the players' scoring probabilities to see how often each team won. As it is not known what the order of penalty takers would have been had the shooutout gone to sudden-death, a random order was selected.

Results

Figure 2 gives the scoring probabilities for all individual players from England and Italy (England in red; Italy in blue) eligible to take a penalty. The England penalty takers selected by manager Southgate were reasonable choices; in fact, owing to his young age, Saka had the highest scoring probability among England players. Italy's choices, on the other hand, appear to be slightly sub-optimal under this model, with players such as Belotti and Bernardeschi appearing towards the bottom of Figure 2.

The outcome from the Monte Carlo simulation was that, given these choices, England would win with probability 0.539 and Italy with probability 0.461. However, there are several limitations that should be pointed out. While the Monte Carlo standard errors are small (the error from using a finite number of simulations), there is uncertainty in the players' scoring probabilities, which is not reflected in this result. This uncertainty could be accounted for by following a Bayesian approach and simulating from the posterior distribution of each p. Also, the simulation assumed that each penalty was independent, but, in practice, players are likely to be influenced by the result of previous penalties.

But England did narrowly come out on top. So, in a sense, it could be argued that England were merely unlucky... better luck in Qatar!

Conclusion

There are, of course, many other factors that influence the outcome of a penalty shootout. Arguably the most important of all is the player's confidence as they place the ball on the penalty spot. Alas, this is perhaps the most difficult of all to quantify. Besides, in the end chance still does play a part, and this is what makes penalties such an exciting - but cruel - way to decide a match or even a tournament.

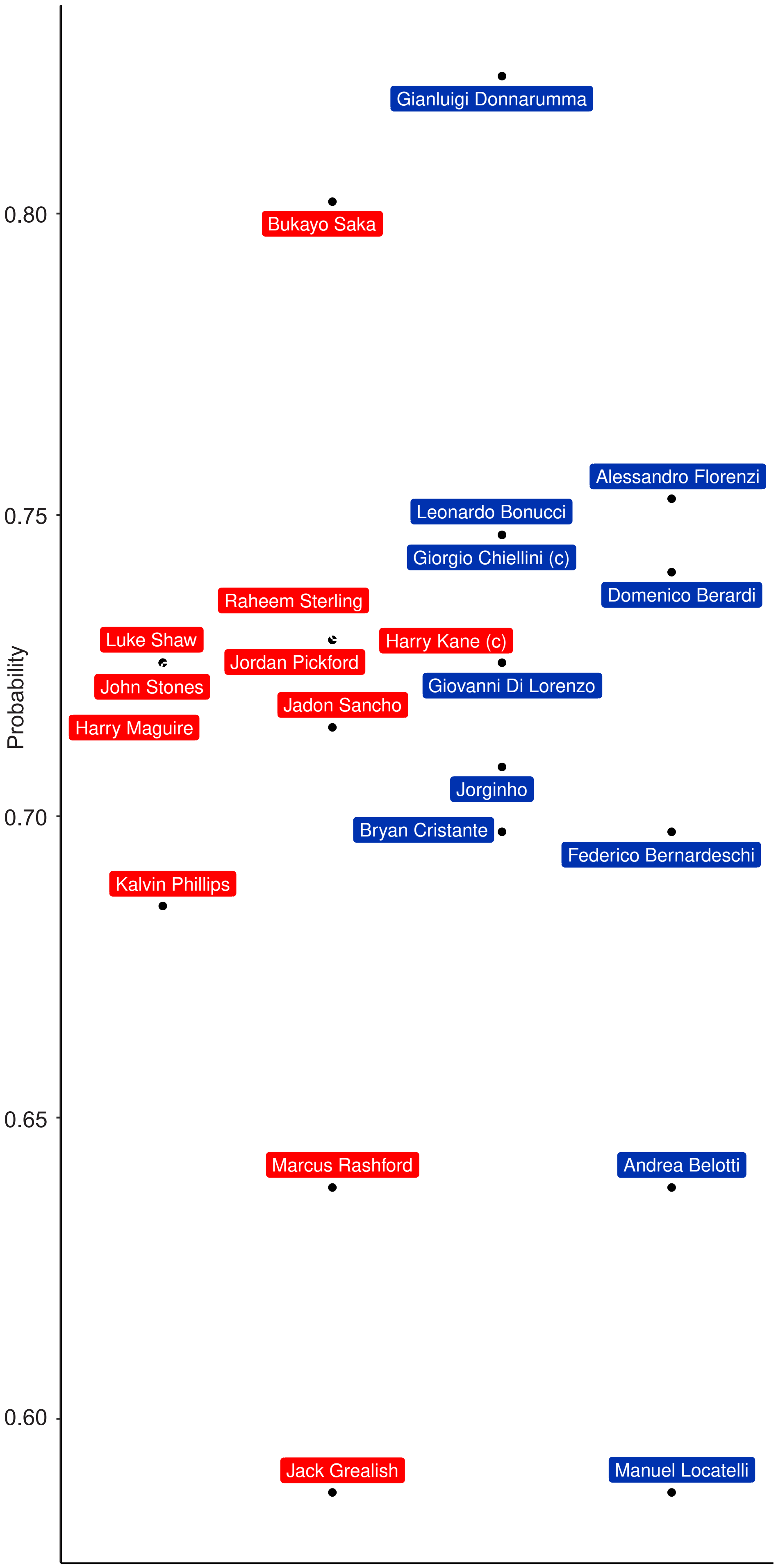

Figure 2: The fitted scoring probabilities for the England and Italy players who were eligible to take a penalty.

References

Jordet, G., Hartman, E., Visscher, C., and Lemmink, K. A. P. M. (2007). Kicks from the penalty mark in soccer: The roles of stress, skill, and fatigue for kick outcomes. *Journal of Sports Sciences*, 25(2):121–129. PMID: 17127587.