

Hanck, Christoph; Arnold, Martin C.

Article — Published Version

Hierarchical Bayes modelling of penalty conversion rates of Bundesliga players

AStA Advances in Statistical Analysis

Provided in Cooperation with:

Springer Nature

Suggested Citation: Hanck, Christoph; Arnold, Martin C. (2021) : Hierarchical Bayes modelling of penalty conversion rates of Bundesliga players, AStA Advances in Statistical Analysis, ISSN 1863-818X, Springer, Berlin, Heidelberg, Vol. 107, Iss. 1-2, pp. 177-204, <https://doi.org/10.1007/s10182-021-00420-w>

This Version is available at:

<https://hdl.handle.net/10419/286781>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Hierarchical Bayes modelling of penalty conversion rates of Bundesliga players

Christoph Hanck¹ · Martin C. Arnold¹

Received: 23 December 2020 / Accepted: 10 September 2021 / Published online: 5 October 2021
© The Author(s) 2021

Abstract

Judging by its significant potential to affect the outcome of a game in one single action, the penalty kick is arguably the most important set piece in football. Scientific studies on how the ability to convert a penalty kick is distributed among professional football players are scarce. In this paper, we consider how to rank penalty takers in the German Bundesliga based on historical data from 1963 to 2021. We use Bayesian models that improve inference on ability measures of individual players by imposing structural assumptions on an associated high-dimensional parameter space. These methods prove useful for our application, coping with the inherent difficulty that many players only take few penalties, making purely frequentist inference rather unreliable.

Keywords Hierarchical Bayes · Shrinkage · Football · Penalties

JEL Classification C11 · C53 · Z20

1 Introduction

Publications from many disciplines deal with the analysis of the world's most popular sport, football. Particular attention is paid to the study of set pieces, especially those that offer imminent opportunity for scoring a goal. For example, Baranda and

Christoph Hanck gratefully acknowledges support by the German Research Foundation (DFG) through the Collaborative Research Center “Statistical modelling of nonlinear dynamic processes” (SFB 823, Project A4). The authors thank the associate editor and two anonymous reviewers for valuable comments which significantly improved the paper.

✉ Martin C. Arnold
martin.arnold@vwl.uni-due.de

Christoph Hanck
christoph.hanck@vwl.uni-due.de

¹ Chair of Econometrics, Faculty of Business Administration and Economics, University of Duisburg-Essen, Universitätsstraße 12, 45117 Essen, Germany

Lopez-Riquelme (2012) and Strafford et al. (2019) study how corner kicks are carried out depending on the current score and examine their relevance for the match outcome. Penalty kicks have also been investigated. Kuhn (1988), Loy (1998), Kamp (2006), and Noël et al. (2014) examine penalty shootouts from psychological and cognitive perspectives based on field experiments and devise strategies for penalty takers. This interest is hardly surprising: for instance, since 1978 about every fifth knockout game at the FIFA World Cup has been decided by penalty shootouts. This results in an estimated chance of 67% that a future world champion must successfully complete at least one penalty shootout on his way to the title (Memmert and Noël 2020).

Penalty kicks in German leagues have been investigated by, e.g. Bornkamp et al. (2009) who use a nonparametric Bayesian approach to rank Bundesliga keepers by penalty saving abilities based on historical data. They find little support for significant differences between goalkeepers, but motivate an investigation of penalty taker ability. Using logistic regression, Kuss et al. (2007) refute the widespread myth that a fouled player should not take the penalty kick because of the alleged greater risk of failure. They argue that this may be due to a selection process that results in penalty shooters being largely homogeneous with respect to their resilience in stress situations. Furthermore, coaches presumably select penalty takers already before the game due to knowledge of such special characteristics and skill.

Historical data such as the ranking of Bundesliga penalty shooters published by the German football magazine Kicker (2021) suggest the existence of significant differences in players' ability to score penalties. However, such rankings are often based on the absolute number of penalties converted and thus fail to capture how efficient players are in exploiting chances. Even rankings based on relative frequencies are questionable: for players frequently nominated as penalty takers, the relative frequency may be a viable estimate. However, for players with small n_i , the actual conversion probability is likely to be estimated imprecisely.

In this paper, we discuss shrinkage estimates from Bayesian beta-binomial models for ranking penalty takers in the German Bundesliga. This approach offers statistical power in estimating the skill of players with little information in borrowing information from players who have taken many penalties. We analyse historical data on 993 players who took a total of 4955 penalties in first division matches from season 1963/64, when the Bundesliga was founded, to 2020/21. We have compiled the necessary dataset from German websites kicker.de and transfermarkt.de. Section 2 examines empirical Bayes estimates from beta-binomial models. We discuss moment matching estimates and estimates based on informative priors obtained from frequentist regression. Section 3 presents more flexible fully Bayesian hierarchical models and compares findings with the results of Sect. 2 using a Bayesian information criterion and out of sample analysis. Based on the outcomes, we compile rankings for active players in the German Bundesliga.¹ Section 4 concludes.

¹ For all models considered in Sects. 2 and 3, we use data up to and including season 2019/20. Penalties taken in season 2020/21 serve as a test dataset for model comparison.

2 Empirical Bayes estimates of penalty conversion ability

The empirical distribution of penalties taken is shown in the left panel of Fig. 1. It is evident that the distribution is right skewed with mode at 1. The empirical distribution of estimated success rates $\hat{\theta}_i := y_i/n_i$, shown in the right panel of Fig. 1, illustrates why rankings based on such a measure are problematic: we find a multimodal distribution with much probability mass near 0 and 1. This is due to many players with few (often a single) attempts who have successfully converted all kicks or, at the other extreme, did not score at all. Owing to insufficient information, the skill of these players is thus likely to be estimated imprecisely by purely frequentist methods.

A preliminary approach for obtaining more reliable estimates of penalty conversion abilities $\theta_i \in [0, 1]$ for a set of players indexed by $i = 1, \dots, N$ is to use a shrinkage estimator in a compound model

$$y_i \sim \pi(y_i|\theta_i), \quad \theta_i \sim \pi(\theta), \quad (1)$$

where θ_i are latent realisations from the ability distribution $\pi(\theta)$. For modelling success rates in sports, this concept was introduced by Efron and Morris (1972, 1977) who analysed baseball hit ratios using the James–Stein estimator which they motivate as a Bayesian point estimator assuming Gaussian $\pi(\cdot)$.

Particularly, for samples with many small n_i it seems more adequate to assume a beta-binomial model, where

$$\begin{aligned} y_i &\sim \text{Bin}(\theta_i, n_i), \\ \theta_i|\alpha, \beta &\sim \text{iid Beta}(\alpha, \beta), \end{aligned} \quad (2)$$

i.e. a player's score y_i in n_i attempts is generated from a binomial distribution with success rate θ_i . We model θ_i via a reparameterised beta prior with expectation $\mu := \frac{\alpha}{\alpha+\beta}$ and precision measure $\nu := \alpha + \beta$. We thus have

$$\pi(\theta_i|y_i) \propto \pi(y_i|\theta_i)\pi(\theta_i|\mu, \nu) \quad (3)$$

$$\propto \theta_i^{y_i+\nu\mu-1}(1-\theta_i)^{n_i-y_i+\nu(1-\mu)-1}, \quad (4)$$

where $\pi(y_i|\theta_i)$ is the usual binomial likelihood and $\pi(\theta_i|\mu, \nu)$ denotes the prior. By conjugacy, the posterior $\pi(\theta_i|y_i)$ is $\text{Beta}(y_i + \nu\mu, n_i - y_i + \nu(1 - \mu))$. An empirical Bayes (EB) approach to conduct inference via (4) specifies an informative prior from point estimates for μ and ν .

2.1 Empirical Bayes estimates using moment matching

An instructive approach discussed by, e.g. Carlin and Louis (2000) uses moment matching (MM) to estimate the hyperparameters. When the n_i are equal across samples, MM estimators are directly derived from the expected value and variance

formulae for a marginal distribution of binomial proportions with success probability $\theta_i \sim \text{Beta}(\nu\mu, \nu(1-\mu))$. The case of unequal n_i is more complicated, as in the compound model (2) the $\hat{\theta}_i$ are independently distributed with $E(\hat{\theta}_i) = \mu$, but $\text{Var}(\hat{\theta}_i) = \mu(1-\mu)n_i^{-1} + (\nu+1)^{-1}\mu(1-\mu)(1-n_i^{-1})$ is a function of n_i . This makes it preferable to give different weight to the $\hat{\theta}_i$. For known variances, Kleinman (1973) suggests to estimate μ by the (best linear unbiased) estimator

$$\hat{\mu} := \frac{\sum_{i=1}^N w_i \hat{\theta}_i}{w}, \quad w := \sum_{i=1}^N w_i \quad (5)$$

with inverse variance weights w_i . The precision ν is then estimated by

$$\hat{\nu} := \frac{\hat{\mu}(1-\hat{\mu}) \left[\sum_{i=1}^N w_i \left(1 - \frac{w_i}{w}\right) - \sum_{i=1}^N \frac{w_i}{n_i} \left(1 - \frac{w_i}{w}\right) \right]}{\hat{\sigma}^2 - \hat{\mu}(1-\hat{\mu}) \left[\sum_{i=1}^N \frac{w_i}{n_i} \left(1 - \frac{w_i}{w}\right) \right]} - 1, \quad (6)$$

which obtains from matching $\hat{\sigma}^2 := \sum_{i=1}^N w_i (\hat{\theta}_i - \hat{\mu})^2$ with the variance of a $\text{Beta}(\nu\mu, \nu(1-\mu))$ random variable.² If w_i are unknown, $\hat{\mu}$ and $\hat{\nu}$ can be computed iteratively using (5) and (6) with initial values $w_i = n_i$ and updates $w_i = n_i / [1 + (\hat{\nu} + 1)^{-1}(n_i - 1)]$, where $\hat{\nu}$ is the current value of $\hat{\nu}$.

Using (4) and estimates $\hat{\mu}$ and $\hat{\nu}$, frequentist conversion probability estimates $\hat{\theta}_i := y_i/n_i$ are then updated by computing posterior means

$$\check{\theta}_i := \hat{S}_i \hat{\mu} + (1 - \hat{S}_i) \hat{\theta}_i, \quad \hat{S}_i := \frac{\hat{\nu}}{\hat{\nu} + n_i} \quad (7)$$

where \hat{S}_i is the shrinkage factor towards the beta prior for player i . It is evident from (7) that more information on individual ability (that is, large n_i) lessens the effect of the prior on $\check{\theta}_i$, i.e. information from the population of penalty takers has less weight for individual estimates. At the other extreme, the prior information dominates estimates of conversion probabilities when n_i is small, resulting in an adjustment towards the overall mean.

Applied to our data, we obtain hyperparameter estimates $\hat{\mu} = .6788$, and $\hat{\nu} = 1.3405$, resulting in a $\text{Beta}(.9099, .4306)$ prior. The corresponding bimodal density is shown in red in Fig. 1. Posterior quantities are presented in Fig. 2. The left panel of Fig. 2 compares relative frequencies against MM estimates from the corresponding posterior distribution. It can be seen that including prior information has a plausible effect: shrinkage is low for observations where the data provides much information whereas point estimates for small n_i are markedly adjusted. This is especially relevant for observations with $\hat{\theta}_i \in \{0, 1\}$ and small n_i where estimates are most strongly corrected towards $\hat{\mu}$. This is sensible as

² Note that our parameterisation differs from Kleinman (1973) who uses $\mu := \alpha/(\alpha + \beta)$ and $r := (\alpha + \beta + 1)^{-1}$, i.e. $\theta \sim \text{Beta}((r^{-1} - 1)\mu, (r^{-1} - 1)(1 - \mu))$.

Table 1 Score-based all-time top and bottom penalty takers in Bundesliga

RK	Player	y_i	n_i	$\hat{\theta}_i$	RK	Player	y_i	n_i	$\hat{\theta}_i$
1	Manfred Kaltz	53	60	.883	:	:	:	:	:
2	Gerd Müller	51	63	.810	804	Volker Graul	0	1	0
3	Michael Zorc	49	57	.860	804	Volker Ohling	0	1	0
4	Horst-Dieter Hötterges	40	48	.833	804	Waldemar Steubing	0	1	0
5	Robert Lewandowski	35	39	.897	804	Wilfried Kohlars	0	1	0
6	Wilfried Hannes	31	38	.816	804	Wilfried Trenkel	0	1	0
6	Paul Breitner	31	37	.838	804	Willibald Weiss	0	1	0
6	Georg Volkert	31	35	.886	804	Wulf-Ingo Usbeck	0	1	0
9	Lothar Matthäus	30	33	.909	804	Yildiray Bastürk	0	1	0
10	Günter Netzer	29	35	.829	804	Zoran Mamic	0	1	0
:	:	:	:	:	804	Zvonimir Soldo	0	1	0

Data for seasons 1963/64–2020/21 from Kicker (2021). RK is the score-based rank. Due to equal scores, the lower ranks are regularly shared by many players which we do not list completely for brevity. y_i is the number of scores. n_i denotes the number of penalties taken. $\hat{\theta}_i := y_i/n_i$

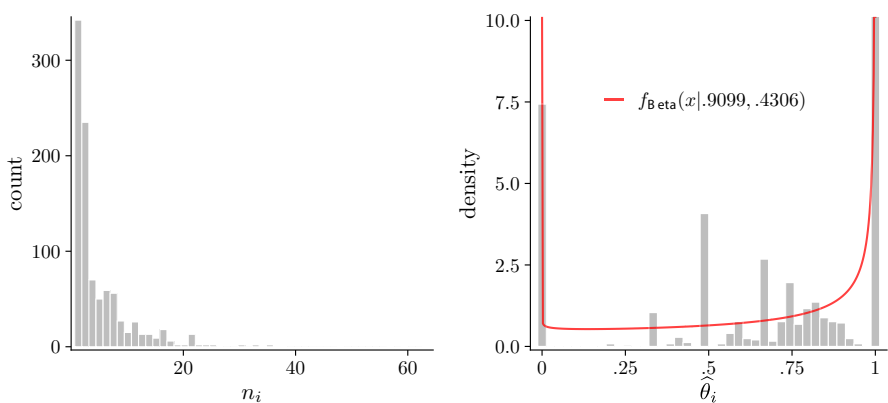


Fig. 1 Historic distribution of penalties taken and penalty conversion rates. *Notes:* Data from Kicker (2021). Left: absolute frequencies of penalties taken in seasons 1963/64–2019/20. Right: distribution of penalty conversion rates along with the corresponding MM-based beta prior density (red line) (colour figure online)

accurate estimates are possible for players with many penalties whereas players with only a few penalties are shrunk towards the mean.

Furthermore, the posterior distributions shown in the right panel of Fig. 2 show that for players with a small n_i , dominance of the prior over the likelihood leads to greater uncertainty of the estimated conversion probability. The effect of shrinkage can also be judged from the dashed black line which shows a kernel density estimate (KDE) of the distribution of posterior means. The shape of the KDE indicates that MM gives more realistic estimates of expected scoring

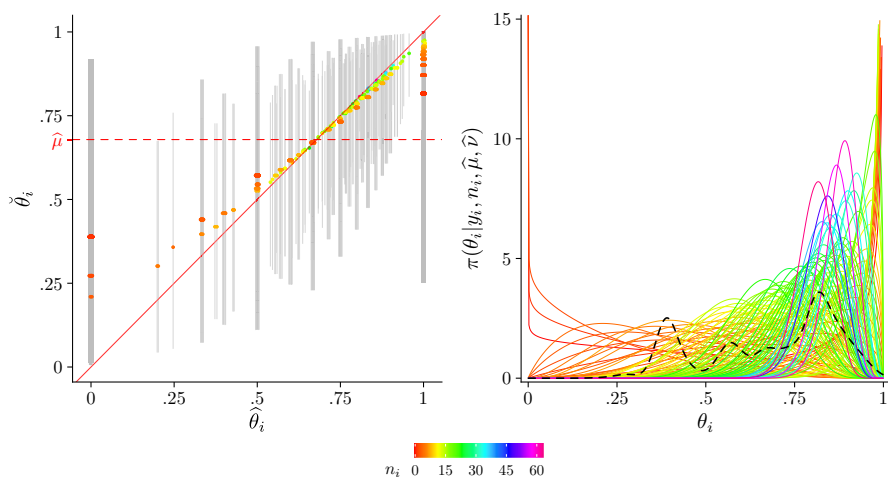


Fig. 2 Empirical Bayes—shrinkage towards a beta prior from moment matching. *Notes:* Penalties taken in seasons 1963/64–2019/20. Left: relative frequencies $\hat{\theta}_i$ and posterior means $\hat{\theta}_i$ with horizontal $\mathcal{U}(-.005, .005)$ jitter. The solid red line marks the set of equal estimates. The dashed red line shows the prior mean estimate $\hat{\mu} = .67$. Grey bars show 95% credible intervals. Right: Beta posterior distributions. The dashed black line shows a kernel density estimate of the distribution of posterior means (colour figure online)

probabilities than the pure frequentist estimates. This is especially evident for the tails of the distribution which have substantially less probability mass compared to the density of the $\hat{\theta}_i$ in Fig. 1.

Table 2 lists the top and bottom 10 historical ranks of Bundesliga penalty kickers based on MM estimates along with 95% credible intervals (CIs). Interestingly, compared to the score-based ranking in Table 1 we find that none of the top listed players made it into the best 10, which are now led by Hans-Joachim Abel with 100% scoring frequency in 16 trials. However, the other top players also converted all their penalties, i.e. the correction due to the prior distribution is small and score probability estimates and CIs are consistently close to 1. For the last 10 ranks, we find that shrinkage leads to somewhat stronger adjustments of the $\hat{\theta}_i$. Tim Borowski, a former German international, for example, is attested a posterior mean scoring probability of 21% despite 3 failed penalties out of 3 attempts. However, this estimate is also subject to considerable uncertainty as his 95% CI is [.005, .643].

2.2 Empirical Bayes estimates from informative beta priors

Several points should be emphasised when interpreting the estimates from the MM approach in Sect. 2.1. First, the prior has most probability mass at extreme abilities and high variance, i.e. MM causes the prior weight to be strongly shaped by players for whom frequentist ability estimates $\hat{\theta}_i$ are most unreliable. Therefore, few penalties are needed for the data to dominate the prior. However, adequacy of the implied shrinkage is an empirical issue and relates to how fast the $\hat{\theta}_i$

Table 2 Empirical Bayes estimates—top and bottom 10 penalty takers

RK	Player	y_i	n_i	$\hat{\theta}_i$	95% CI	RK	Player	y_i	n_i	$\hat{\theta}_i$	95% CI
32	Hans-Joachim Abel	16	16	.975	[.870, 1]	:	:	:	:	:	:
37	Ludwig Nolden	15	15	.974	[.862, 1]	791	Kai Michalke	0	2	.272	[.007, .768]
43	Max Kruse	14	14	.972	[.854, 1]	791	Lars Hermel	0	2	.272	[.007, .768]
60	Otto Rehhagel	12	12	.968	[.832, 1]	791	Markus Elmer	0	2	.272	[.007, .768]
72	Mehmet Scholl	11	11	.965	[.819, 1]	791	Michael Tarnat	0	2	.272	[.007, .768]
97	Alfred Finnbogason	9	9	.958	[.786, 1]	791	Norbert Meier	0	2	.272	[.007, .768]
97	Horst Wild	9	9	.958	[.786, 1]	791	Sigfried Held	0	2	.272	[.007, .768]
97	Márcio Amoroso	9	9	.958	[.786, 1]	791	Sven Demandt	0	2	.272	[.007, .768]
116	Heinz Flohe	8	8	.954	[.765, 1]	791	Uwe Klimaschefski	0	2	.272	[.007, .768]
116	Manfred Bender	8	8	.954	[.765, 1]	791	Dieter Krafczyk	0	3	.210	[.005, .643]
:	:	:	:	:	:	791	Tim Borowski	0	3	.210	[.005, .643]

Penalties taken in seasons 1963/64–2019/20. RK is the score-based rank. Due to equal scores, the lower ranks are regularly shared by many players which we do not list completely for brevity

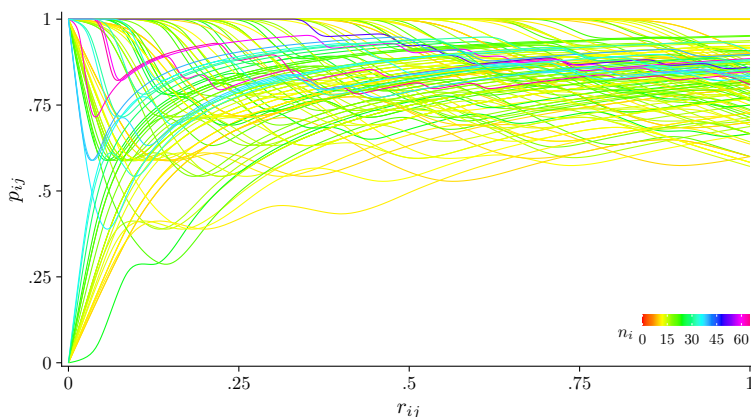


Fig. 3 Convergence of relative frequencies. *Note:* Trajectories of relative frequency estimates for $n_i \geq 10$. $r_{ij} := (j-1)/(n_i-1)$, $p_{ij} := \sum_{l=1}^j y_{il}/j$, $j \in \{1, \dots, n_i\}$ with $y_{il} \in \{0, 1\}$ the outcome of penalty l for player i . Spline fits are shown for better readability

converge towards the true abilities as players progressively take penalties. To get an impression of the convergence speed Fig. 3 visualises trajectories of the $\hat{\theta}_i$ over time for a sub-sample of the players: as ability estimates inherently stabilise later in the career, we consider players with $n_i \geq 10$. Even for players with $n_i \approx 20$, we observe enough time-variability in scoring rates for shrinkage towards the prior to be attractive. Notably, given the MM estimate $\hat{v} = 1.3405$ it can be computed from (7) that the shrinkage factor \hat{S}_i decays quickly with n_i and is already at .4 for only $n_i = 2$ attempts, which seems rather fast.

Figure 3 also shows evidence of the selection process outlined in Sect. 1: we do not observe players who repeatedly fail to score and yet continue taking penalties: players in the sub-sample who start off with a failed attempt usually have a scoring streak afterwards and thus keep getting nominated. This selection might result in overestimation of abilities for players who have performed poorly and only had a few attempts.

We next consider EB estimates based on informative beta priors

$$\theta_i | \hat{\mu}_i^s, \hat{v}^s \sim \text{Beta}(\hat{v}^s \hat{\mu}_i^s, \hat{v}^s (1 - \hat{\mu}_i^s)). \quad (8)$$

For this, we estimate prior hyperparameters incorporating player performance data \mathbf{x}_i by maximising the frequentist regression beta-binomial likelihood

$$L(\boldsymbol{\gamma}, \sigma_0 | \mathbf{y}) = \prod_{i=1}^N f(y_i | \boldsymbol{\gamma}, \sigma_0), \quad (9)$$

$$f(y_i | \boldsymbol{\gamma}, \sigma_0) = \binom{n_i}{y_i} \frac{\Gamma(y_i + \alpha_i^s) \Gamma(n_i - y_i + \beta_i^s) \Gamma(\alpha_i^s + \beta_i^s)}{\Gamma(n_i + \alpha_i^s + \beta_i^s) \Gamma(\alpha_i^s) \Gamma(\beta_i^s)}, \quad (10)$$

with $\Gamma(\cdot)$ the Gamma function. Prior mean $\mu_i^s := \alpha_i^s / v^s$ and precision $v^s := \alpha_i^s + \beta_i^s$ have logit and log links in linear predictors,

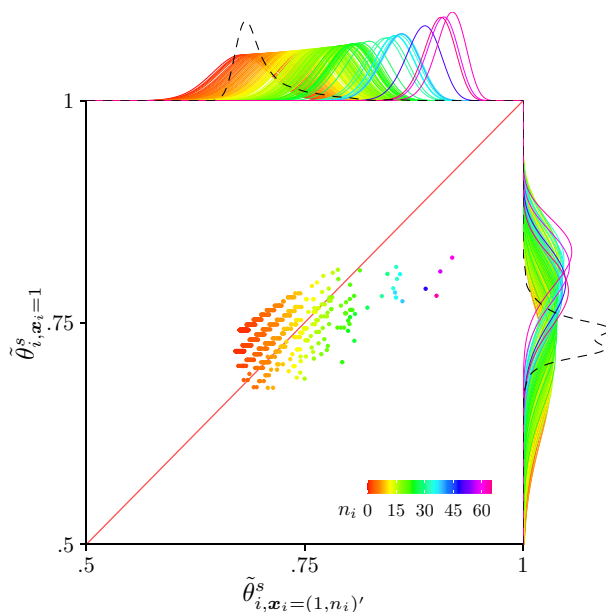


Fig. 4 Likelihood-based empirical Bayes—comparison of posterior quantities. *Note:* Dots show posterior means based on priors obtained from beta-binomial regression on x_i , with horizontal $\mathcal{U}(-.005, .005)$ jitter. Beta posterior distributions are displayed at the margins. The dashed black lines show kernel density estimates of posterior means

$$\mu_i^s := \text{logit}^{-1}(x_i' \gamma), \quad v^s := \exp(\sigma_0)^{-1}. \quad (11)$$

To get a notion of the selection effect and the impact of likelihood-based prior parameter estimates, we compare posterior means from two preliminary models first: one where the μ_i^s are assumed constant across players and one where μ_i^s depend on n_i . Note that players who are known as good penalty kickers will take penalties more often, making n_i a natural predictor of the scoring ability.

For the baseline model, we obtain hyperparameter estimates $\hat{\mu}^s = .74$ and $\hat{v}^s = 41.04$. The latter can be interpreted as the prior sample size and gives the threshold beyond which the data dominates the prior in contributing to posterior means, see (7). This suggests that the impact of the prior is significantly stronger than for MM estimates, cf. Fig. 2. Figure 4 shows posterior distributions and means for both preliminary models and provides evidence of stronger shrinkage. This is particularly noticeable for players with extreme scoring ratios and few attempts, although players with large n_i are now also subject to stronger adjustment towards the prior. The red dots above the 45 deg-line show that players with few attempts tend to be rated worse in the model adjusted for n_i than in the baseline model with only a constant. Consequently, frequently nominated players tend to be better rated once we accommodate for the effect of selection: the colour transition in posterior means indicates that penalty takers are systematically better rated from approximately 20 penalties onwards as priors become more

Table 3 Summary statistics of Bundesliga penalty takers

	Description	Min	Q25	Q50	Q75	Max	Mean	SD
<i>n</i>	Penalty attempts	1	1	2	6	63	4.96	6.51
<i>y</i>	Successes	0	1	2	4.25	53	3.71	5.50
<i>saved</i>	Saved penalties	0	0	1	1	12	.91	1.18
<i>miss</i>	Missed penalties	0	0	0	1	5	.34	.63

Penalties taken in seasons 1963/64–2019/20. Q25 to Q75 denote quartiles of the empirical distribution

informative. This is also reflected in estimation uncertainty displayed by posterior distributions where the variance is generally lower for the adjusted model and decreases markedly with n_i , cf. the densities shown above the square. We discuss ability point estimates, the implied shrinkage, and resulting rank reversals with respect to results from MM in more detail for fully Bayesian estimation in Sect. 3.1.

We also consider priors that additionally depend on *miss*, the number of missed penalties, i.e. penalty kicks that did not require activity by the goalkeeper because the ball went past the goal, hit the crossbar or one of the posts, see Table 3 for summary statistics on the regressors. The rationale behind this is as follows: players whose penalties are often harmless for the keeper are presumably bad candidates and are also less likely to be nominated again. Our dataset contains 265 players with at least one penalty that did not require intervention by the goalkeeper. In total, 54 of these players were not nominated again after missing their first and only penalty. Including *miss* as a predictor for the scoring probability in $\mathbf{x}_i'\boldsymbol{\gamma}$ should attenuate the correction towards the population average for these players and also allow better discrimination between players with similar $\hat{\theta}_i$ but different failure rate. Frequentist coefficient estimates $\hat{\boldsymbol{\gamma}}$ are given in Table 8 in [Appendix]. We find evidence for the expected effects, i.e. significant positive (negative) estimates for *n* (*miss*).

Posterior distributions and their means for informative priors are shown in Fig. 5. As anticipated from the results for the basic models in Fig. 3, the shrinkage in the extended model is significantly stronger than for MM. The KDE for the posterior means indicates that point estimates are significantly concentrated here as well and only occasionally fall below 50% (the median is .6978). Nevertheless, our prior choice yields substantial differences to relative frequency estimates in some cases. For example, former HSV player Sergej Barbarez with four missed penalties in nine attempts is now the worst rated player with a posterior mean of .426 (cf. Table 4 and the leftmost orange-coloured posterior distribution in Fig. 4). Though below the average posterior mean of 69.7%, Tim Borowski (2 saves and 1 miss in 3 attempts) is rated much better now with a posterior scoring probability of 60.7% (and 95% CI [.503, .706]). At the other extreme, prior information yields more optimistic and less uncertain estimates for frequently nominated and thus potentially better players. The ranking is led by Michael Zorc whose posterior mean of 92% exceeds his score ratio of 86%, cf. Table 1.

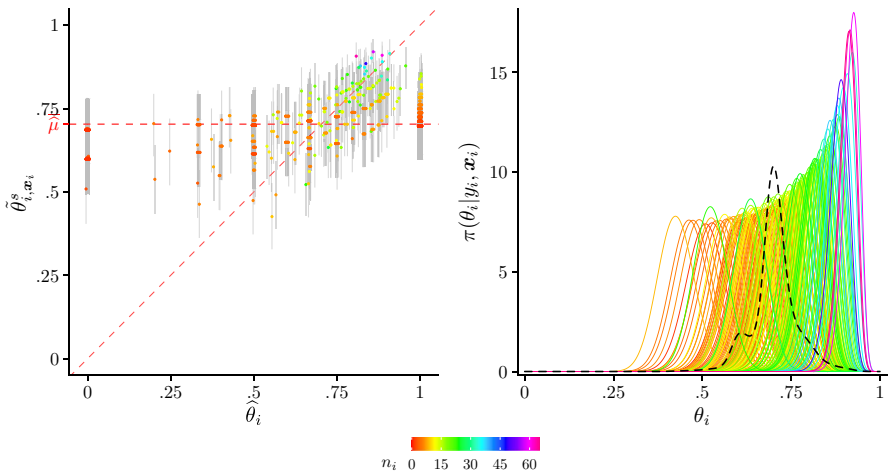


Fig. 5 Empirical Bayes estimates from informative beta priors-posterior quantities. *Notes:* $\mathbf{x}_i = (1, n_i, \text{miss}_i)'$. Left: the solid red line marks the set of equal relative frequencies $\hat{\theta}_i$ and posterior means $\hat{\theta}_{i,x}^s$. The dashed red line shows the prior mean estimate $\hat{\mu} = .7033$. Grey bars show 95% credible intervals. Right: beta posterior distributions. The dashed black line shows a kernel density estimate of the distribution of posterior means (colour figure online)

3 Fully Bayesian analysis of penalty conversion probabilities

A potential drawback of the methods discussed in Sect. 2 is the choice of prior: we obtain inference on the θ_i using individual as well as prior information which depends on aggregate information from all observations through hyperparameter point estimates. However, using plug-in estimators for μ and ν (or α and β) implies using an *estimated* posterior distribution. Doing so bears the risk of overfitting and being overly optimistic about the precision of the results. This is especially problematic for players with small n_i . A refined approach hence is to assume a joint probability model for all parameters: instead of letting prior information reflect a (conditional) population average, we allow information on the individual conversion abilities to contribute probabilistically to the posterior of the hyperparameters, which then shapes posteriors of the θ_i . This is achieved in fully Bayesian hierarchical models where we put priors (hyperpriors) on the model hyperparameters.

3.1 A fully Bayesian beta-binomial hierarchical model

For the baseline hierarchical model (HB) we retain model distributions as in (2) but obtain the joint posterior $\pi(\theta, \alpha, \beta | \mathbf{y})$ from a fully Bayesian approach that assumes an exchangeable beta prior, that is, $\pi(\theta | \alpha, \beta) = \prod_{i=1}^N \pi(\theta_i | \alpha, \beta)$. Together with the binomial likelihood $\pi(\mathbf{y} | \theta, \alpha, \beta)$ and a hyperprior $\pi(\alpha, \beta)$ we thus have

$$\pi(\theta, \alpha, \beta | \mathbf{y}) \propto \pi(\mathbf{y} | \theta, \alpha, \beta) \pi(\theta | \alpha, \beta) \pi(\alpha, \beta) \quad (12)$$

Table 4 Subjective Empirical Bayes estimates—top and bottom penalty takers

RK	Player	y_i	$miss_i$	n_i	$\hat{\theta}_{ix_i}^s$	95% CI	RK	Player	y_i	$miss_i$	n_i	$\hat{\theta}_{ix_i}^s$	95% CI
3	Michael Zorc	49	1	57	.920	[.870, .959]	:	:	:	:	:	:	:
8	Lothar Matthäus	30	0	33	.915	[.859, .958]	791	Tim Borowski	0	1	3	.607	[.503, .706]
1	Manfred Kaltz	53	3	60	.910	[.858, .951]	:	:	:	:	:	:	:
2	Gerd Müller	51	0	63	.907	[.856, .948]	523	Marko Pantelic	1	2	5	.538	[.435, .640]
9	Stefan Kuntz	29	0	34	.902	[.842, .948]	523	Hartmut Heidemann	1	2	3	.526	[.422, .630]
5	Georg Volkert	31	1	35	.890	[.829, .940]	170	Heinz Strehl	6	3	9	.525	[.424, .625]
4	Horst-Dieter Höttinges	40	2	48	.885	[.826, .933]	37	Jürgen Grabowski	15	5	23	.522	[.428, .615]
15	Rainer Bonhof	20	0	21	.878	[.810, .933]	791	Uwe Klimaschefski	0	2	2	.509	[.404, .613]
5	Wilfried Hannes	31	1	38	.878	[.815, .929]	245	Rudolf Brunnenmeier	4	3	7	.491	[.390, .593]
9	Günter Netzer	29	1	35	.874	[.809, .927]	305	Michael Frontzeck	3	3	6	.474	[.373, .577]
:	:	:	:	:	:	:	397	Harry Decheteir	2	3	6	.463	[.362, .566]
:	:	:	:	:	:	:	203	Sergej Barbarez	5	4	9	.426	[.328, .527]

Penalties taken in seasons 1963/64–2019/20. $\mathbf{x}_i = (1, n_i, miss_i)'$. RK is the score-based rank

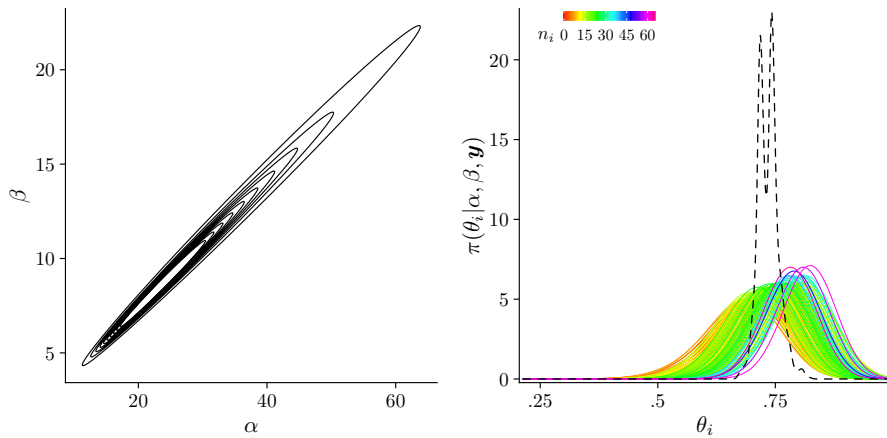


Fig. 6 Fully Bayesian hierarchical model-simulated posterior distributions. *Note:* Left: unnormalised marginal posterior $\pi(\alpha, \beta|y)$ computed over $(\alpha, \beta) \in [0, 80] \times [0, 30]$. Contour lines indicate 5%, 15%, ..., 85%, 95% of the density at the mode (23.42, 8.87). Right: Kernel estimates of posterior densities $\pi(\theta_i|\alpha, \beta, y)$

$$\propto \prod_{i=1}^N \theta_i^{y_i} (1 - \theta_i)^{n_i - y_i} \prod_{i=1}^N \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1 - \theta_i)^{\beta-1} \pi(\alpha, \beta). \quad (13)$$

By standard results, the marginal posterior of (α, β) follows as

$$\pi(\alpha, \beta|y) \propto \prod_{i=1}^N \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_i)\Gamma(\beta + n_i - y_i)}{\Gamma(\alpha + \beta + n_i)} \pi(\alpha, \beta). \quad (14)$$

Regarding the choice of an “objective” hyperprior we follow Gelman et al. (2013) and use the Pareto kernel

$$\pi(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}, \quad (15)$$

which obtains by placing a uniform density on $\varphi(\alpha, \beta) := (\mu, \nu^{-1/2})$ and using the Jacobian of φ to transform back to the original scale.³ Showing that using (15) as a prior for (α, β) in (14) yields a proper marginal posterior if $0 < y_i < n_i$ for at least one i follows from standard arguments.⁴

Contours of $\pi(\alpha, \beta|y)$ based on a numeric evaluation of the functional in (14) with $\pi(\alpha, \beta)$ as in (15) are shown in the left panel of Fig. 6. The marginal posterior density has a mode at (23.42, 8.87) and mean at (33.01, 11.83) which, chosen as hyperparameters in a beta prior, roughly agrees with expected value and prior sample size

³ Being a precision measure of the beta prior, ν inversely affects the variance of the scoring probabilities. Placing a uniform prior on $\nu^{-1/2}$ is thus approximately equivalent to placing a diffuse prior on the standard deviation of θ .

⁴ See Gelman et al. (2013) for further discussion of prior choice.

of the beta prior in the baseline likelihood-based EB model of Sect. 2.⁵ This seems reasonable since both models rely on the full likelihood but $\pi(\alpha, \beta)$ in HB introduces additional uncertainty. We thus expect that shrinkage towards the prior will be stronger than for MM.

The left panel of Fig. 6 reveals two marked features of $\pi(\alpha, \beta|\mathbf{y})$. First, the contours indicate that regions with $\alpha + \beta \leq 2$ have virtually zero density. Second, the correlation between α and β is strongly positive. By properties of the beta distribution, the marginal posterior thus places nearly all probability mass on hyperparameter combinations (α, β) that yield unimodal beta priors with similar central tendency. This is markedly different from the bimodal prior obtained from MM in Sect. 2.1 which has most probability mass in the tails, giving substantial weight to extreme outcomes. Furthermore, the additional uncertainty from imposing the hyperprior (15) on (α, β) implies higher uncertainty in outcomes from the (conditional) posterior distribution in the hierarchical model. Consequently, we expect estimates of posterior quantities to have higher variance than MM estimates. We will address these aspects in more detail in Sect. 3.2.

Although the $\pi(\theta_i|\alpha, \beta, \mathbf{y})$ are beta distributions, uncertainty in (α, β) in HB does not allow direct simulation from the conditional posteriors as for the empirical Bayes models where using a standard routine for generating beta variates with fixed hyperparameters suffices. Instead, a procedure that incorporates variability in the hyperparameters due to the hyperprior (15) is needed. The idea is to first generate samples from the marginal posterior $\pi(\alpha, \beta|\mathbf{y})$ which are then used to draw from $\pi(\theta|\alpha, \beta, \mathbf{y})$. We use the following easily implemented discrete-grid sampling scheme suggested in Gelman et al. (2013) to generate random variates from the joint posterior $\pi(\theta, \alpha, \beta|\mathbf{y})$.

Sampling from the joint posterior $\pi(\theta, \alpha, \beta|\mathbf{y})$.

1. *Simulate from $\pi(\alpha, \beta|\mathbf{y})$.* Draw $(\alpha_\ell, \beta_\ell)$, $\ell = 1, \dots, L$ for sufficiently large L using the following discrete-grid sampling procedure:
 - 1.1 Compute the (unnormalised) marginal posterior $\pi(\alpha, \beta|\mathbf{y})$ with hyperprior (15) over a sufficiently large grid. Normalise by approximating the density as a step function with unit integral over the grid.
 - 1.2 Obtain $\pi(\alpha|\mathbf{y})$ by discrete summation.

For $\ell = 1, \dots, L$:

- 1.3 Draw α_ℓ from $\pi(\alpha|\mathbf{y})$ and draw β_ℓ from $\pi(\beta|\alpha, \mathbf{y})$ given α_ℓ using the discrete inverse CDF method.
- 1.4 Add a uniform zero-mean jitter to α_ℓ and β_ℓ to get continuously distributed samples.

⁵ We have also considered alternative weakly informative prior specifications for μ and ν , e.g. independent uniform and gamma. However, this led to virtually the same results which we omit for brevity.

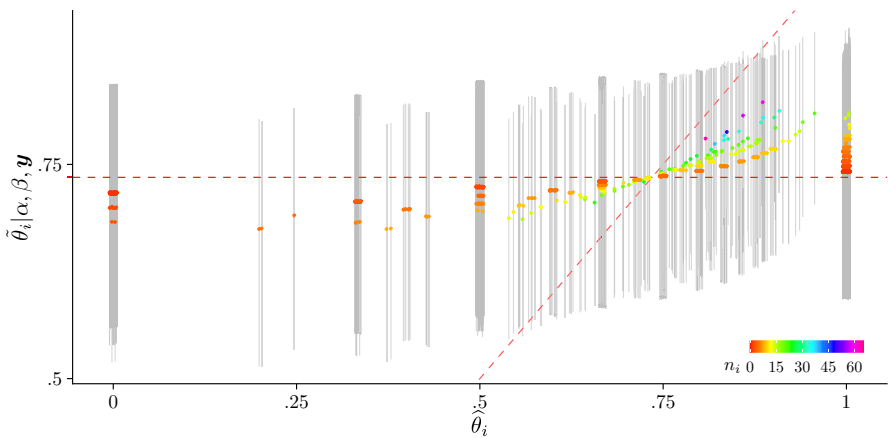


Fig. 7 Fully Bayesian hierarchical model—conditional posterior means. *Note:* the solid red line marks the set of equal relative frequencies $\hat{\theta}_i$ and posterior mean estimates $\hat{\theta}_i | \alpha, \beta, \mathbf{y}$. The dashed red line shows the average posterior mean. The $\hat{\theta}_i$ are jittered to reduce overplotting. Grey bars show 95% credible intervals (colour figure online)

2. *Simulate from $\pi(\theta | \alpha, \beta, \mathbf{y})$.* For each $i = 1, \dots, N$, use the $(\alpha_\ell, \beta_\ell)$ to obtain L draws of θ_i from the conditional posterior distribution $\theta_i | \alpha, \beta, \mathbf{y} \sim \text{Beta}(\alpha + y_i, \beta + n_i - y_i)$.

3.1.1 Results

Posterior quantities are approximated using $L = 50000$ draws generated with the above sampling scheme. The right panel of Fig. 6 shows KDEs of the posterior distributions of the $\theta_i | \alpha, \beta, \mathbf{y}$. We find that modelling uncertainty in the hyperparameters by the joint probability model results in the $\pi(\theta_i | \alpha, \beta, \mathbf{y})$ being significantly shrunk towards each other. As for the EB models, this correction is most pronounced for penalty takers with little information since the beta prior in the hierarchical model is shaped significantly by players with many attempts—a sub-sample whose information is more relevant for inference on players where data is scarce. Comparing the kernel density estimate for the distribution of posterior means to the distribution of MM posterior means in Fig. 2 further illustrates this feature.

Figure 7 compares posterior means and relative frequencies $\hat{\theta}_i$ to examine shrinkage of frequentist estimates implied by the fully Bayesian approach somewhat more closely. Consistent with the discussion of implications of the hyperprior in the previous section we find more pronounced shrinkage on the estimated penalty kicking ability of the players. It is noticeable that modelling the hyperparameters via the joint probability distribution results in a pronounced correction of scoring probabilities towards the posterior population average for players with an intermediate number of attempts and $\hat{\theta}_i \in (0.5, 1)$ —a sub-sample for which MM estimates often differ only marginally from $\hat{\theta}_i$. Two further features distinguish the results from those of the MM approach. First, failed penalties are more influential in HB than for MM, i.e.

the posterior is adjusted towards lower scoring probabilities if a player demonstrates lack of skill in several attempts. Consequently, we expect order reversals in the outcomes relative to MM. Furthermore, for $\hat{\theta}_i$ close to unity the uncertainty as visualised by vertical grey bars in Fig. 7 is larger than for MM where the precision of the results seems to be exaggerated, cf. Fig. 2.

Table 5 shows the top and bottom 10 ranks for posterior mean estimates in the fully Bayesian hierarchical model along with 95% CIs. Overall, we find that the posterior means are corrected more strongly compared to MM. Lower bounds of the 95% CIs for top players fall in a more realistic range near the average posterior scoring probability. We find that the all-time ranking in the hierarchical model is led by Manfred Kaltz, now with Lothar Matthäus in second place. Hans-Joachim Abel holds the third place but is closely followed by several renowned players, amongst them Michael Zorc and the presumably best active penalty taker in first-tier Bundesliga, Bayern Munich's Robert Lewandowski, see also Table 7.⁶

As expected, the upward correction for players in the lower tail of the ability distribution is more pronounced in the hierarchical model. Again, the 95% CIs appear more reasonable, the more so as upper bounds overlap with the intervals of upper-ranking players. This appears natural: given pressure through the public, trainers tend to waive assigning players to further penalties once they have failed some, especially when further candidates are on the team. Yet, these players—being trained professionals—of course would be able to perform with a substantially higher success rate than observed so far if they were assigned to further penalties.

For the bottom ranks, we observe more order reversals which, as explained above, result from higher failure weighting implied by the prior. Tim Borowski, for instance, is again ranked amongst the worst penalty takers in Bundesliga history but ranks slightly better than Bruno Labbadia with only one goal in five attempts and posterior mean conversion probability of about 67.5%.⁷ Interestingly, Borowski ranks almost as high as Klaas-Jan Huntelaar who failed in 7 out of 16 attempts.

In view of these results, it is worth discussing how appropriate the implied shrinkage is. Mean and mode as well as the highest density region in the conditional posterior distribution for (α, β) in Fig. 6 suggests that about 30–40 penalties are needed for the data to dominate the prior. This finding fits well with the maximum likelihood estimate $\hat{\nu} \approx 41$ for the baseline model in Sect. 2.2. However, considering the convergence of relative frequencies reported in Fig. 3 this value may seem too high. That said, model adequacy should not be assessed entirely by discrepancies between Bayesian and frequentist estimates. For a more informed evaluation of model fits Sect. 3.2 compares models using Bayesian information criteria and a test sample covering penalties in season 2020/21.

⁶ We recommend googling Michael Kutzop for the remarkable story of his single missed penalty.

⁷ Note that discrepancies in $\hat{\theta}_i$ and CIs for players with equal performance are an artefact from approximating posterior distributions by simulation.

Table 5 Fully Bayesian hierarchical model—top and bottom 10 penalty takers

RK	Player	y_i	n_i	$\hat{\theta}_i$	95% CI	RK	Player	y_i	n_i	$\hat{\theta}_i$	95% CI
1	Manfred Kaltz	53	60	.823	[.742, .895]	:	:	:	:	:	:
8	Lothar Matthäus	30	33	.813	[.717, .899]	97	Klaas-Jan Huntelaar	9	16	.686	[.549, .798]
32	Hans-Joachim Abel	16	16	.810	[.699, .910]	791	Dieter Krafczyk	0	3	.683	[.519, .811]
15	Rainer Bonhof	20	21	.809	[.703, .904]	397	Harry Decheiver	2	6	.683	[.527, .806]
37	Ludwig Nolden	15	15	.807	[.694, .908]	791	Tim Borowski	0	3	.683	[.520, .811]
3	Michael Zorc	49	57	.807	[.723, .882]	397	Dieter Eckstein	2	6	.682	[.526, .806]
12	Robert Lewandowski	27	30	.805	[.706, .894]	397	Marco Weibhaupt	2	6	.682	[.527, .807]
5	Georg Volkert	31	35	.804	[.709, .890]	523	Marko Pantelic	1	5	.675	[.514, .801]
43	Max Kruse	14	14	.804	[.690, .906]	305	Manfred Pohlschmidt	3	8	.675	[.522, .798]
27	Michael Kutzop	17	18	.799	[.688, .899]	523	Bruno Labbadia	1	5	.675	[.514, .803]
:	:	:	:	:	:	305	Günter Herrmann	3	8	.675	[.520, .797]

Penalties taken in seasons 1963/64–2019/20. RK is the score-based rank. $\hat{\theta}_i$ and 95% CI denote the estimated mean and the 95% credible interval of the respective conditional posterior distribution $\pi(\theta_i | \alpha, \beta, y)$

Table 6 Comparison of Bayesian beta-binomial models

model	ELPD _{LOO}	Δ ELPD _{LOO}	SE _{Δ}	LOOIC	SE _{LOOIC}	<i>ll</i>
EB $\mathbf{x}_i = (1, n_i, \text{miss}_i)'$	− 1121.31	0.00	0.00	2242.63	39.81	0.47
BBBR $\mathbf{x}_i = (1, n_i, \text{miss}_i)'$	− 1125.64	− 4.33	0.65	2251.28	39.87	0.47
EB $\mathbf{x}_i = (1, n_i)'$	− 1159.14	− 37.83	8	2318.28	38.71	0.48
BBBR $\mathbf{x}_i = (1, n_i)'$	− 1162.82	− 41.51	7.46	2325.65	37.98	0.49
EB $\mathbf{x}_i = 1$	− 1201.69	− 80.37	11.01	2403.38	41.90	0.48
BBBR $\mathbf{x}_i = 1$	− 1203.19	− 81.88	11.10	2406.38	42.14	0.48
HB	− 1203.58	− 82.27	10.92	2407.16	41.79	0.48
EB (moment matching, MM)	− 1415.17	− 293.86	17.67	2830.35	50.76	0.52
frequentist						1.34

ELPD_{LOO} is the LOO estimate of the expected log pointwise predictive density. Δ ELPD_{LOO} is the difference in ELPD relative to the best model. SE _{Δ} the standard error of component-wise differences in ELPD. LOOIC = $-2 \cdot \text{ELPD}_{\text{LOO}}$ with standard error SE_{LOOIC}. *ll* is the binary cross-entropy for season 2020/21 predictions. Bayesian models are fitted with the NUTS of Stan (Carpenter et al. 2017) via the brms R package (Bürkner 2017). Posterior quantities are obtained based on 50000 samples generated from four independent Markov chains

3.2 Additional Markov chain Monte Carlo estimates and model comparison

It seems worthwhile—e.g. for predictive purposes—to know which active first division players are the most likely to convert future penalties. To this end, we first assess adequacy of the discussed models. We also consider further models that satisfy the Bayesian paradigm and incorporate uncertainty in hyperparameters. Here, we use the same regressor sets as for the EB models and fit fully Bayesian beta-binomial regressions (BBBR) with logit and identity links for μ and ν , setting vague Gaussian and gamma priors. Posterior quantities for these models are obtained from Markov Chain Monte Carlo (MCMC) sampling.⁸

For Bayesian applications, metrics based on an estimate of the (expected log) predictive density (ELPD) have been suggested for evaluating out-of-sample predictive performance. For comparison of the Bayesian models, we use the leave-one-out cross-validation criterion (LOOIC) proposed by Vehtari et al. (2017) which approximates the LOOCV estimate of ELPD by regularised importance sampling. Results are given in Table 6. Generally, the EB models with informative priors and their MCMC counterparts perform very similarly, yet MCMC fits are reported to be slightly inferior in each case, which is likely due to prior induced uncertainty in parameters. The MM approach is clearly dominated by the other models. As expected, the beta-binomial regression models with constant linear predictor from Sect. 2.2 perform very similarly as HB with Pareto prior for (α, β) .⁹ LOOIC indicates

⁸ In BBBR models, we use $\mathcal{N}(0, 1.5)$ priors for the γ_i and set a $\text{Gamma}(4.5, 0.1)$ prior for ν which, similarly as for the fully Bayesian hierarchical model, has expectation 45 but roughly 95% probability mass on $[1, 85]$.

⁹ Note that the HB model can be cast as a BBBR with linear predictor $\mu = \gamma_0 \sim \mathcal{U}[0, 1]$ and $\nu = \sigma_0 \sim \text{Pareto}(\eta, -1.5)$ with reasonably small lower bound $\eta > 0$.

that the most complex beta-binomial regression model performs best among the fully Bayesian models.¹⁰

For a comparison with the naïve frequentist strategy of using independent binomial proportions $\hat{\theta}_i$ as ability estimates, we compute binary cross-entropy loss using predicted scoring probabilities for 94 penalties kicked in the 2020/21 season by players with $n_i \geq 1$.¹¹ As is seen from the ll column in Table 6, log loss for the test dataset also indicates that MM performs worst among the Bayesian models, while full BBBR does best. Clearly, the frequentist approach performs worst.

We now discuss the results for the BBBR model with $\mathbf{x}_i = (1, n_i, miss_i)'$ in more detail and examine ability estimates of top and bottom active Bundesliga players.¹² The results are shown in Table 7.¹³ The ranking is topped by Robert Lewandowski with a posterior scoring probability of 88.4%, closely followed by Max Kruse and Andrej Kramaric.¹⁴ At the bottom of the first division ranking, we find Borussia Mönchengladbach's Breel Embolo and Sebastian Rudy of TSG Hoffenheim who both missed their only Bundesliga penalties to date. Again, we conclude that rank reversals from the BBBR model with $\mathbf{x}_i = (1, n_i, miss_i)'$ can be insightful. For example, Erling Haaland (one save and one miss in four attempts) with all-time score-based rank 406 is rated almost as Thorgan Hazard (2 saves and 2 misses out of 12 attempts) and all-time rank 119.

4 Discussion and outlook

This article conducts an analysis of the penalty-kicking abilities of Bundesliga players for a historical sample covering seasons 1963/64 to 2020/21. An inherent problem in estimating scoring probabilities and establishing a ranking is that the data are imbalanced among players. From a statistical perspective, sparsity in observations resulting from many penalty takers having only taken a few penalties is challenging. Purely likelihood-based estimation where players are treated as independent then yields imprecise results.

We have shown that a ranking based on hierarchical Bayesian models in which the accuracy of all players is modelled by a joint probability distribution have advantages over a purely frequentist approach. In comparison to common rankings based on the relative scoring frequency or even the absolute number of goals, such as that of Kicker (2021), we find some notable rank reversals using posterior estimates from

¹⁰ Following a suggestion of an anonymous referee, it would also be interesting to jointly model the (categorical) probability distribution of the triplet $(y, miss, save)$ instead of using counts of misses as a predictor for the scoring probability in BBBR, in order to account for differing *rates* instead of *counts* of misses. We however leave this idea for future research for brevity here.

¹¹ For ll to be defined under the frequentist paradigm, we let $\hat{\theta}_i = 1 - 10^{-5}$ to handle cases where $y_i = n_i$.

¹² Posterior quantities for model parameters are shown in Table 9 in [Appendix].

¹³ Complete results for active first division players are given in Table 10 of [Appendix], which also gives player performance data at the end of season 2020/21. We have also estimated conversion probabilities for historical data on 2nd tier Bundesliga. See Table 12 of [Appendix] for the emerging ranking of active players.

¹⁴ Estimates based on a dataset that also includes DFB Cup penalties put Union Berlin's Max Kruse ahead of Robert Lewandowski. However, credible intervals have considerable overlap, see Table 11 in [Appendix].

Table 7 BBBR—top and bottom active penalty takers in Bundesliga

RK	player	y_i	$miss_i$	n_i	$\hat{\theta}_i$	95% CI	RK	player	y_i	$miss_i$	n_i	$\hat{\theta}_i$	95% CI
5	Robert Lewandowski	35	2	39	.884	[.812, .941]	:	:	:	:	:	:	:
18	Max Kruse	19	0	20	.878	[.800, .942]	119	Thorgan Hazard	8	2	12	.645	[.531, .751]
15	Andrej Kramaric	20	1	22	.839	[.754, .912]	311	Ondrej Duda	3	1	4	.644	[.523, .756]
72	Nils Petersen	11	0	14	.808	[.712, .890]	406	Erling Haaland	2	1	4	.630	[.507, .742]
34	Daniel Caligiuri	16	1	19	.803	[.710, .883]	535	Janik Haberer	1	1	2	.611	[.484, .727]
119	Vincenzo Grifo	8	0	8	.790	[.689, .879]	535	Charles Aranguiz	1	1	2	.611	[.485, .726]
119	André Silva	8	0	8	.789	[.689, .879]	535	Admir Mehmedi	1	1	2	.610	[.484, .727]
119	Daniel Brosinski	8	0	8	.789	[.688, .879]	535	Adam Szalai	1	1	2	.610	[.484, .726]
45	Thomas Müller	14	1	16	.787	[.691, .872]	174	Dodi Lukebakio	6	2	8	.610	[.491, .724]
119	Nabil Bentaleb	8	0	9	.786	[.685, .874]	804	Breel Embolo	0	1	1	.593	[.464, .711]
:	:	:	:	:	:	:	804	Sebastian Rudy	0	1	1	.592	[.464, .711]

Data from Kicker (2021). RK is the score-based all-time rank. $\hat{\theta}_i$ and 95% CI denote the posterior mean and the 95% CI of the respective conditional posterior distribution

hierarchical models. This predominantly concerns lower quantiles of the distribution where the shrinkage is greatest. Changes in the top ranks are usually small. This is as expected, as players in the top ranks include those that tend to convert many penalties. Hence, they get assigned to further penalties, resulting in many data points for those players. Thus, purely likelihood-based inference already is fairly informative.

Although the Bayesian results may be statistically more convincing than a comparison of plain scoring frequencies, the posterior mean results should be interpreted with consideration of the associated uncertainty. While differences in posterior means are not negligible—suggesting that there are indeed players who are significantly more accurate on penalties than others—credible intervals tend to overlap considerably. This pattern is evident both for the historical overall ranking and also for separate rankings of active players in the first and second Bundesliga.

For further research, it might be interesting to use a dataset that provides more detailed information on the individual penalty kicks for each player to develop a model that can be used for predicting outcomes of penalties in future Bundesliga games.

Appendix: Additional results

See Tables 8, 9, 10, 11, 12.

Table 8 Frequentist beta-binomial regression estimates

	Estimate	SD	<i>t</i> value	Pr(> <i>t</i>)
$\hat{\mu}$				
Intercept	.77	.05	14.45	< 0.001
<i>n</i>	.05	.00	12.49	< 0.001
<i>miss</i>	– .39	.04	– 9.12	< 0.001
$\hat{\sigma}_0$				
Intercept	– 4.44	.28	– 16.03	< 0.001

Data covering seasons 1963/64–2019/20 from Kicker (2021)

Table 9 Bayesian beta-binomial regression: posterior quantities

	Estimate	Post. SE	l-95% CI	u-95% CI	\hat{R}
μ					
Intercept	.76	.06	.65	.87	1
<i>n</i>	.05	.01	.04	.06	1
<i>miss</i>	– .40	.05	– .49	– .31	1
ν					
Intercept	67.78	18	39.90	109.22	1

Data covering seasons 1963/64–2020/21 from Kicker (2021). Post. SE is the estimated standard deviation of the posterior distribution. l-95% CI and u-95% CI are lower and upper bounds of the 95% credible interval. \hat{R} is the Gelman–Rubin statistic

Table 10 Bayesian beta-binomial regression estimates—active tier 1 Bundesliga players

RK	Player	n_i	y_i	saved	miss	$\hat{\theta}_i$	$\tilde{\theta}_i$	95% CI
5	Robert Lewandowski	39	35	2	2	.897	.884	[.812, .941]
18	Max Kruse	20	19	1	0	.950	.878	[.800, .942]
15	Andrej Kramaric	22	20	1	1	.909	.839	[.754, .912]
72	Nils Petersen	14	11	3	0	.786	.808	[.712, .890]
34	Daniel Caligiuri	19	16	2	1	.842	.803	[.710, .883]
119	Vincenzo Grifo	8	8	0	0	1	.790	[.689, .879]
119	André Silva	8	8	0	0	1	.789	[.689, .879]
119	Daniel Brosinski	8	8	0	0	1	.789	[.688, .879]
45	Thomas Müller	16	14	1	1	.875	.787	[.691, .872]
119	Nabil Bentaleb	9	8	1	0	.889	.786	[.685, .874]
155	Eric Maxim Choupo-Moting	9	7	2	0	.778	.772	[.669, .862]
52	Marco Reus	17	13	3	1	.765	.771	[.673, .857]
174	Niclas Füllkrug	7	6	1	0	.857	.763	[.658, .856]
208	Vladimir Darida	6	5	1	0	.833	.752	[.643, .847]
208	Jonas Hofmann	6	5	1	0	.833	.751	[.642, .847]
88	Wout Weghorst	11	10	0	1	.909	.744	[.642, .839]
251	Marcel Sabitzer	4	4	0	0	1	.740	[.629, .840]
251	Lucas Alario	5	4	1	0	.80	.739	[.628, .838]
251	Milot Rashica	5	4	1	0	.80	.738	[.628, .835]
98	Emil Forsberg	11	9	1	1	.818	.731	[.625, .827]
311	Marcel Halstenberg	3	3	0	0	1	.726	[.612, .828]
311	Silas Katompa Mvumpa	3	3	0	0	1	.726	[.612, .829]
311	Ramy Bensebaini	3	3	0	0	1	.726	[.611, .829]
311	Nicolás González	4	3	1	0	.750	.725	[.612, .826]
98	Klaas-Jan Huntelaar	17	9	7	1	.529	.722	[.615, .815]
406	Daniel Ginczek	2	2	0	0	1	.712	[.594, .818]
406	Daniel Didavi	2	2	0	0	1	.712	[.595, .817]
406	Michael Gregoritsch	2	2	0	0	1	.712	[.594, .817]
406	Amine Harit	2	2	0	0	1	.712	[.596, .816]
406	Davie Selke	2	2	0	0	1	.712	[.595, .817]
406	Robin Quaison	2	2	0	0	1	.711	[.594, .817]
406	Krzysztof Piatek	3	2	1	0	.667	.711	[.594, .815]
406	Sebastian Andersson	3	2	1	0	.667	.711	[.594, .815]
406	Matheus Cunha	3	2	1	0	.667	.711	[.594, .814]
535	Leon Goretzka	1	1	0	0	1	.697	[.577, .804]
535	Steven Zuber	2	1	1	0	.50	.697	[.577, .803]
535	Ihlas Bebou	1	1	0	0	1	.697	[.578, .806]
535	Serge Gnabry	2	1	1	0	.50	.697	[.578, .804]
535	Jadon Sancho	1	1	0	0	1	.697	[.575, .805]
535	Roland Sallai	1	1	0	0	1	.697	[.576, .804]
535	Luka Jovic	1	1	0	0	1	.697	[.577, .805]
535	Dani Olmo	1	1	0	0	1	.697	[.576, .804]
535	Leonardo Bittencourt	1	1	0	0	1	.697	[.577, .805]
535	Florian Kainz	1	1	0	0	1	.697	[.575, .805]
535	Makoto Hasebe	2	1	1	0	.50	.697	[.577, .803]

Table 10 (continued)

RK	Player	n_i	y_i	saved	miss	$\hat{\theta}_i$	$\tilde{\theta}_i$	95% CI
535	Gonzalo Castro	1	1	0	0	1	.697	[.577, .805]
535	Lars Bender	1	1	0	0	1	.697	[.577, .805]
535	Mark Uth	3	1	2	0	.333	.697	[.577, .801]
535	Kerem Demirbay	1	1	0	0	1	.696	[.576, .805]
535	Christian Gentner	2	1	1	0	.50	.696	[.577, .803]
535	Moussa Niakhaté	1	1	0	0	1	.696	[.576, .804]
535	Gonçalo Paciência	1	1	0	0	1	.696	[.577, .804]
535	Marcus Ingvartsen	2	1	1	0	.50	.696	[.577, .803]
535	Fabian Klos	2	1	1	0	.50	.696	[.576, .801]
535	Filip Kostic	3	1	2	0	.333	.696	[.576, .800]
535	André Hahn	3	1	2	0	.333	.696	[.577, .802]
804	Munas Dabbur	1	0	1	0	0	.682	[.558, .791]
804	Lucas Höler	1	0	1	0	0	.681	[.558, .790]
804	Jonathan Schmid	1	0	1	0	0	.681	[.558, .790]
804	Anthony Ujah	1	0	1	0	0	.681	[.558, .791]
804	Ömer Toprak	1	0	1	0	0	.681	[.557, .790]
804	Christian Günter	1	0	1	0	0	.681	[.558, .789]
804	Philipp Förster	1	0	1	0	0	.681	[.559, .791]
804	Steven Skrzybski	1	0	1	0	0	.681	[.559, .791]
804	Maximilian Philipp	1	0	1	0	0	.681	[.558, .790]
804	Ermin Bicakcic	1	0	1	0	0	.681	[.558, .790]
804	Alassane Plea	1	0	1	0	0	.681	[.559, .789]
804	Alexander Sörloth	1	0	1	0	0	.681	[.558, .789]
804	Maximilian Arnold	1	0	1	0	0	.681	[.557, .790]
804	Mathew Leckie	1	0	1	0	0	.681	[.558, .789]
251	Florian Niederlechner	6	4	1	1	.667	.662	[.546, .769]
98	Lars Stindl	12	9	1	2	.750	.658	[.546, .764]
98	Alfred Finnbogason	11	9	0	2	.818	.657	[.544, .764]
311	David Alaba	5	3	1	1	.60	.646	[.527, .755]
311	Wendell	5	3	1	1	.60	.646	[.526, .755]
119	Thorgan Hazard	12	8	2	2	.667	.645	[.531, .751]
311	Ondrej Duda	4	3	0	1	.750	.644	[.523, .756]
406	Erling Haaland	4	2	1	1	.50	.630	[.507, .742]
535	Janik Haberer	2	1	0	1	.50	.611	[.484, .727]
535	Charles Aránguiz	2	1	0	1	.50	.611	[.485, .726]
535	Admir Mehmedi	2	1	0	1	.50	.610	[.484, .727]
535	Ádám Szalai	2	1	0	1	.50	.610	[.484, .726]
174	Dodi Lukebakio	8	6	0	2	.750	.610	[.491, .724]
804	Breel Embolo	1	0	0	1	0	.593	[.464, .711]
804	Sebastian Rudy	1	0	0	1	0	.592	[.464, .711]

Historic Bundesliga data from Kicker (2021). RK is the score-based all-time rank in tier 1 Bundesliga

Table 11 Bayesian beta-binomial regression estimates—active tier 1 Bundesliga players (extended data-set)

RK	Player	n_i	y_i	saved	miss	$\hat{\theta}_i$	$\tilde{\theta}_i$	95% CI
18	Max Kruse	24	23	1	0	.950	.892	[.821, .950]
5	Robert Lewandowski	42	37	3	2	.897	.880	[.811, .935]
15	Andrej Kramaric	23	21	1	1	.909	.840	[.758, .910]
72	Nils Petersen	17	14	3	0	.786	.828	[.740, .901]
34	Daniel Caligiuri	21	18	2	1	.842	.814	[.728, .889]
119	André Silva	10	10	0	0	1	.807	[.713, .890]
119	Nabil Bentaleb	11	10	1	0	.889	.803	[.710, .885]
119	Vincenzo Grifo	9	9	0	0	1	.797	[.702, .882]
119	Daniel Brosinski	9	9	0	0	1	.797	[.702, .882]
45	Thomas Müller	24	20	2	2	.875	.781	[.691, .862]
52	Marco Reus	18	14	3	1	.765	.778	[.685, .860]
174	Niclas Füllkrug	8	7	1	0	.857	.774	[.674, .862]
155	Eric Maxim Choupo-Moting	9	7	2	0	.778	.771	[.670, .859]
251	Marcel Sabitzer	5	5	0	0	1	.755	[.649, .849]
208	Jonas Hofmann	6	5	1	0	.833	.753	[.647, .846]
251	Lucas Alario	6	5	1	0	.80	.753	[.648, .845]
208	Vladimir Darida	7	5	2	0	.833	.751	[.647, .842]
98	Emil Forsberg	12	10	1	1	.818	.743	[.643, .834]
311	Marcel Halstenberg	4	4	0	0	1	.743	[.633, .838]
251	Milot Rashica	5	4	1	0	.80	.741	[.635, .836]
311	Nicolás González	5	4	1	0	.750	.741	[.634, .836]
406	Daniel Ginczek	3	3	0	0	1	.731	[.621, .829]
311	Ramy Bensebaini	3	3	0	0	1	.731	[.621, .827]
311	Silas Katomba Mvumpa	3	3	0	0	1	.730	[.622, .829]
98	Klaas-Jan Huntelaar	17	9	7	1	.529	.723	[.619, .812]
406	Amine Harit	2	2	0	0	1	.718	[.605, .819]
406	Davie Selke	2	2	0	0	1	.718	[.607, .818]
535	Jadon Sancho	2	2	0	0	1	.717	[.607, .817]
406	Daniel Didavi	2	2	0	0	1	.717	[.605, .818]
406	Robin Quaison	2	2	0	0	1	.717	[.605, .819]
406	Michael Gregoritsch	2	2	0	0	1	.717	[.604, .819]
535	Lars Bender	2	2	0	0	1	.717	[.606, .818]
406	Matheus Cunha	3	2	1	0	.667	.717	[.605, .816]
406	Krzysztof Piatek	3	2	1	0	.667	.717	[.605, .816]
406	Sebastian Andersson	3	2	1	0	.667	.716	[.604, .815]
535	André Hahn	4	2	2	0	.333	.715	[.603, .812]
535	Florian Kainz	1	1	0	0	1	.704	[.590, .806]
535	Gonçalo Paciência	1	1	0	0	1	.704	[.590, .807]
535	Dani Olmo	1	1	0	0	1	.704	[.589, .807]
535	Moussa Niakhaté	1	1	0	0	1	.704	[.589, .807]
535	Leonardo Bittencourt	1	1	0	0	1	.704	[.589, .806]
535	Roland Sallai	1	1	0	0	1	.704	[.589, .808]
535	Leon Goretzka	1	1	0	0	1	.704	[.590, .807]
535	Luka Jovic	1	1	0	0	1	.704	[.588, .807]

Table 11 (continued)

RK	Player	n_i	y_i	saved	miss	$\hat{\theta}_i$	$\tilde{\theta}_i$	95% CI
535	Gonzalo Castro	1	1	0	0	1	.704	[.589, .807]
535	Kerem Demirbay	1	1	0	0	1	.704	[.588, .808]
535	Ihlas Bebou	1	1	0	0	1	.704	[.588, .808]
535	Steven Zuber	2	1	1	0	.50	.704	[.590, .805]
535	Fabian Klos	2	1	1	0	.50	.703	[.590, .804]
535	Marcus Ingvartsen	2	1	1	0	.50	.703	[.590, .805]
804	Philipp Förster	2	1	1	0	0	.703	[.588, .806]
535	Serge Gnabry	2	1	1	0	.50	.703	[.589, .805]
535	Makoto Hasebe	2	1	1	0	.50	.703	[.589, .806]
535	Filip Kostic	3	1	2	0	.333	.703	[.588, .803]
535	Christian Gentner	3	1	2	0	.50	.702	[.588, .803]
535	Mark Uth	4	1	3	0	.333	.701	[.587, .801]
119	Thorgan Hazard	15	11	2	2	.667	.690	[.583, .785]
804	Munas Dabbur	1	0	1	0	0	.690	[.572, .794]
804	Maximilian Arnold	1	0	1	0	0	.690	[.572, .793]
804	Alassane Plea	1	0	1	0	0	.690	[.573, .794]
804	Christian Günter	1	0	1	0	0	.690	[.572, .794]
804	Lucas Höler	1	0	1	0	0	.690	[.571, .794]
804	Ermin Bicakcic	1	0	1	0	0	.690	[.571, .793]
804	Maximilian Philipp	1	0	1	0	0	.690	[.573, .794]
804	Ömer Toprak	1	0	1	0	0	.690	[.572, .793]
804	Steven Skrzybski	1	0	1	0	0	.689	[.572, .794]
804	Mathew Leckie	1	0	1	0	0	.689	[.572, .793]
804	Anthony Ujah	1	0	1	0	0	.689	[.571, .793]
804	Jonathan Schmid	1	0	1	0	0	.689	[.572, .794]
804	Alexander Sörloth	1	0	1	0	0	.689	[.573, .793]
98	Lars Stindl	13	10	1	2	.750	.677	[.570, .776]
88	Wout Weghorst	13	10	1	2	.909	.677	[.571, .777]
98	Alfred Finnbogason	12	10	0	2	.818	.677	[.570, .778]
311	Wendell	6	4	1	1	.60	.670	[.557, .772]
251	Florian Niederlechner	6	4	1	1	.667	.670	[.557, .773]
311	Ondrej Duda	5	4	0	1	.750	.670	[.555, .774]
311	David Alaba	5	3	1	1	.60	.656	[.540, .760]
406	Erling Haaland	4	2	1	1	.50	.641	[.523, .749]
535	Admir Mehmedi	2	1	0	1	.50	.625	[.504, .736]
535	Àdám Szalai	2	1	0	1	.50	.625	[.504, .736]
535	Charles Aranguiz	2	1	0	1	.50	.624	[.504, .736]
535	Janik Haberer	2	1	0	1	.50	.624	[.504, .735]
174	Dodi Lukebakio	8	6	0	2	.750	.621	[.508, .730]
804	Breel Embolo	1	0	0	1	0	.608	[.486, .722]
804	Sebastian Rudy	1	0	0	1	0	.608	[.484, .722]

Historic Bundesliga and DFB Cup data from Kicker (2021). RK is the score-based all-time rank in tier 1 Bundesliga

Table 12 Bayesian beta-binomial regression estimates—active tier 2 Bundesliga players

RK	Player	n_i	y_i	saved	miss	$\hat{\theta}_i$	$\tilde{\theta}_i$	95% CI
8	Dimitrij Nazarov	21	20	1	0	.952	.976	[.927, .998]
44	Alexander Mühling	14	14	0	0	1	.972	[.920, .998]
74	Rouwen Hennings	11	10	1	0	.909	.939	[.866, .985]
110	Kevin Wolze	9	8	1	0	.889	.924	[.841, .978]
89	Kevin Behrens	11	9	2	0	.818	.923	[.840, .976]
147	Kingsley Schindler	7	7	0	0	1	.922	[.839, .978]
110	Marvin Ducksch	10	8	2	0	.80	.915	[.829, .972]
21	Tobias Kempe	20	17	2	1	.850	.912	[.831, .967]
25	Marc Schnatterer	18	16	1	1	.889	.911	[.830, .968]
16	Simon Terodde	25	18	7	0	.720	.90	[.808, .960]
181	Dennis Srbeny	8	6	2	0	.750	.897	[.802, .961]
234	Danny Blum	5	5	0	0	1	.897	[.803, .964]
234	Marvin Knoll	5	5	0	0	1	.897	[.804, .964]
234	Branimir Hrgota	6	5	1	0	.833	.892	[.796, .960]
234	Silvere Ganvoula	6	5	1	0	.833	.892	[.797, .959]
234	Nick Proschwitz	6	5	1	0	.833	.892	[.796, .959]
234	Aaron Hunt	6	5	1	0	.833	.892	[.796, .959]
288	Pascal Testroet	4	4	0	0	1	.881	[.782, .956]
288	Christoph Moritz	4	4	0	0	1	.881	[.782, .955]
288	Max Besuschkow	6	4	2	0	.667	.873	[.768, .947]
376	Dominik Kaiser	3	3	0	0	1	.864	[.759, .945]
376	Dawid Kownacki	3	3	0	0	1	.864	[.758, .945]
376	Johannes Geis	3	3	0	0	1	.864	[.757, .945]
376	Andreas Albers	3	3	0	0	1	.864	[.758, .944]
376	Manuel Schäffler	4	3	1	0	.750	.863	[.757, .941]
376	Julian Green	4	3	1	0	.750	.862	[.756, .941]
376	Marvin Wanitzek	4	3	1	0	.750	.862	[.755, .941]
288	Philipp Hofmann	8	4	4	0	.50	.862	[.751, .939]
376	Sebastian Kerk	5	3	2	0	.60	.860	[.751, .938]
512	Tim Kleindienst	2	2	0	0	1	.845	[.732, .932]
512	Sonny Kittel	2	2	0	0	1	.845	[.731, .932]
512	Rodrigo Zalazar	3	2	1	0	.667	.845	[.732, .930]
512	Stefan Maierhofer	2	2	0	0	1	.844	[.733, .932]
512	Janni Serra	3	2	1	0	.667	.844	[.731, .930]
512	Marvin Pieringer	2	2	0	0	1	.844	[.732, .932]
512	Guido Burgstaller	3	2	1	0	.667	.844	[.731, .929]
512	Daniel Keita-Ruel	4	2	2	0	.50	.844	[.730, .928]
700	Sebastian Stolze	2	1	1	0	.50	.825	[.706, .916]
700	Denis Thomalla	2	1	1	0	.50	.824	[.705, .917]
700	Jann George	2	1	1	0	.50	.824	[.705, .917]
700	Leart Paqarada	2	1	1	0	.50	.824	[.704, .916]
700	Patrick Schmidt	2	1	1	0	.50	.824	[.705, .917]

Table 12 (continued)

RK	Player	n_i	y_i	saved	miss	$\hat{\theta}_i$	$\tilde{\theta}_i$	95% CI
700	Paul Seguin	2	1	1	0	.50	.824	[.705, .916]
700	Christian Kühlwetter	2	1	1	0	.50	.824	[.704, .916]
700	Chris Führich	1	1	0	0	1	.823	[.704, .917]
700	Tobias Mohr	1	1	0	0	1	.823	[.704, .917]
700	Dominic Baumann	1	1	0	0	1	.822	[.703, .918]
700	David Pisot	1	1	0	0	1	.822	[.703, .918]
700	Martin Kobylanski	1	1	0	0	1	.822	[.702, .917]
700	Niklas Schmidt	1	1	0	0	1	.822	[.702, .918]
700	Fabian Holland	1	1	0	0	1	.822	[.704, .917]
700	Christian Santos	1	1	0	0	1	.822	[.702, .917]
700	Alexander Esswein	1	1	0	0	1	.822	[.703, .918]
700	Mitja Lotric	1	1	0	0	1	.822	[.701, .917]
700	Oliver Hein	1	1	0	0	1	.822	[.702, .917]
700	Omar Marmoush	1	1	0	0	1	.822	[.703, .918]
700	Genki Haraguchi	1	1	0	0	1	.822	[.701, .917]
700	Ken Reichel	1	1	0	0	1	.822	[.702, .918]
700	Lukas Gugganig	1	1	0	0	1	.822	[.703, .917]
700	Benjamin Kessel	1	1	0	0	1	.822	[.703, .917]
700	Hanno Behrens	1	1	0	0	1	.822	[.704, .916]
700	Arne Feick	1	1	0	0	1	.822	[.702, .917]
1129	Andreas Geipl	2	0	2	0	0	.804	[.679, .901]
1129	Robert Herrmann	2	0	2	0	0	.804	[.677, .901]
1129	Manuel Riemann	1	0	1	0	0	.802	[.677, .900]
1129	Tim Kister	1	0	1	0	0	.802	[.676, .901]
1129	Dong-Won Ji	1	0	1	0	0	.802	[.674, .900]
1129	Marc Lorenz	1	0	1	0	0	.802	[.674, .901]
1129	Sebastian Schonlau	1	0	1	0	0	.801	[.676, .900]
1129	Christian Strohdiek	1	0	1	0	0	.801	[.674, .900]
1129	Felix Lohkemper	1	0	1	0	0	.801	[.674, .900]
1129	Jan Hochscheidt	1	0	1	0	0	.801	[.673, .901]
234	Serdar Dursun	6	5	0	1	.833	.707	[.577, .823]
288	Robert Zulj	5	4	0	1	.80	.675	[.541, .797]
512	Enrico Valentini	4	2	1	1	.50	.621	[.479, .749]
512	Robert Leipertz	3	2	0	1	.667	.602	[.460, .736]
1129	Felix Platte	1	0	0	1	0	.520	[.370, .665]

Historic tier 2 Bundesliga from Kicker (2021). RK is the score-based all-time rank in tier 2 Bundesliga

Acknowledgements We thank Janine Langerbein and Jens Klenke for excellent research assistance.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baranda, P.S.D., Lopez-Riquelme, D.: Analysis of corner kicks in relation to match status in the 2006 world cup. *Eur. J. Sport Sci.* **12**(2), 121–129 (2012)
- Bornkamp, B., Fritsch, A., Kuss, O., Ickstadt, K.: Penalty specialists among goalkeepers: a nonparametric Bayesian analysis of 44 years of German Bundesliga, pp. 63–76. Physica-Verlag HD, Heidelberg (2009)
- Bürkner, P.C.: Brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**(1), 1–28 (2017)
- Carlin, B.P., Louis, T.A.: Bayes and empirical Bayes methods for data analysis, 2nd edn. Chapman & Hall, London (2000)
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: a probabilistic programming language. *J. Stat. Softw.* **76**(1), (2017)
- Efron, B., Morris, C.: Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika* **59**(2), 335–347 (1972)
- Efron, B., Morris, C.: Stein's paradox in statistics. *Sci. Am.* **236**, 119–127 (1977)
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D.: Bayesian Data Analysis, 3rd edn. Chapman & Hall, London (2013)
- Kamp, J.V.D.: A field simulation study of the effectiveness of penalty kick strategies in soccer: late alterations of kick direction increase errors and reduce accuracy. *J. Sports Sci.* **24**(5), 467–477 (2006)
- Kicker: Elfmeterstatistik der Spieler: Bundesliga. (2021). <https://www.kicker.de/bundesliga/elfmeter-schuetzen-historie>. Accessed 10 June 2021
- Kleinman, J.C.: Proportions with extraneous variance: single and independent samples. *J. Am. Stat. Assoc.* **68**(341), 46–54 (1973)
- Kuhn, W.: Penalty-kick strategies for shooters and goalkeepers. In Reilly, T. (ed). *Science and Football: Proceedings of the First World Congress of Science and Football*, Liverpool, 13–17th April 1987. E. & F.N. Spon, London (1988)
- Kuss, O., Kluttig, A., Stoll, O.: The fouled player should not take the penalty himself: an empirical investigation of an old German football myth. *J. Sports Sci.* **25**(9), 963–967 (2007)
- Loy, R.: Handlungsstrategien von Torhütern und Schützen in der Strafstoßsituation des Fussballsports. *Sportwissenschaft rund um den Fussball*, pp. 67–78 (1998)
- Memmert, D., Noël, B.: The Penalty Kick: The Psychology of Success. Meyer & Meyer Sport, London (2020)
- Noël, B., Furley, P., van der Kamp, J., Dicks, M., Memmert, D.: The development of a method for identifying penalty kick strategies in association football. *J. Sports Sci.* **1–10** (2014)
- Strafford, B.W., Smith, A., North, J.S., Stone, J.A.: Comparative analysis of the top six and bottom six teams' corner kick strategies in the 2015/2016 English premier league. *Int. J. Perform. Anal. Sport* **19**(6), 904–918 (2019)
- Vehtari, A., Gelman, A., Gabry, J.: Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**(5), 1413–1432 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.