Methods for Paired Comparisons with Ties

Valerie Bradley, Emmanuelle Dankwa, Natalia Garcia Martin

University of Oxford
University of Warwick

28 November 2018

Outline

1. Classical Models and Results
2. Bayesian Models and Results
3. Model Comparisons

Outline

1. **Classical Models and Results**
2. Bayesian Models and Results
3. Model Comparisons

Bradley-Terry model

$$P(i \text{ beats } j) = \frac{\alpha_i}{\alpha_i + \alpha_j}$$

$$L(\boldsymbol{\alpha}) = \prod_{i=1}^{K} \prod_{j=1}^{K} \left(\frac{\alpha_i}{\alpha_i + \alpha_j} \right)^{w_{ij}}$$

- K teams
- Strengths $\alpha_1, \ldots, \alpha_K$
- Odds α_i / α_j
- w_{ij} number of times i beats j $(w_{ii} = 0)$
- Log strengths $\lambda_i = \log \alpha_i$
- $\text{logit}\left[P(i \text{ beats } j)\right] = \log\left(\frac{P(i \text{ beats } j)}{1 - P(i \text{ beats } j)}\right) = \log\left(\frac{P(i \text{ beats } j)}{P(j \text{ beats } i)}\right) = \lambda_i - \lambda_j$

Davidson-Beaver model

- For i playing home:

$$P(i \text{ beats } j) = \frac{\theta\alpha_i}{\theta\alpha_i + \alpha_j + \delta\sqrt{\alpha_i\alpha_j}}$$
$$P(j \text{ beats } i) = \frac{\alpha_j}{\theta\alpha_i + \alpha_j + \delta\sqrt{\alpha_i\alpha_j}}$$
$$P(\text{draw}) = \frac{\delta\sqrt{\alpha_i\alpha_j}}{\theta\alpha_i + \alpha_j + \delta\sqrt{\alpha_i\alpha_j}}$$

- Home advantage parameter $\theta \geq 1$ ($\theta = 1$ for Davidson)
- Draw parameter $\delta \geq 0$ ($\delta = 0$ for B-T model)
- Power $1/2$ vs. $1/3$

Rao-Kupper model (with order effect)

- For i playing home:

$$P(i \text{ beats } j) = \frac{\theta\alpha_i}{\theta\alpha_i + \gamma\alpha_j}$$

$$P(j \text{ beats } i) = \frac{\alpha_j}{\theta\gamma\alpha_i + \alpha_j}$$

$$P(\text{draw}) = \frac{(\gamma^2 - 1)\theta\alpha_i\alpha_j}{(\theta\alpha_i + \gamma\alpha_j)(\theta\gamma\alpha_i + \alpha_j)}$$

- Rao Kupper: $\theta = 1$
- Draw parameter $\gamma \geq 1$ ($\gamma = 1$ for B-T model)

Poisson trick

- `gnm(formula = count` ~ -1 `+ X + home + draw, eliminate = match, family = poisson, data = season.data)`
- Identifiability constraint: $\lambda_{wh} = 0$ $(\alpha_{wh} = 1)$
- $\exp(\lambda_i - \lambda_j) = \alpha_i/\alpha_j$

Home team	Away team	Final outcome	Match	Count
Brighton	Fulham	A	1	0
Brighton	Fulham	D	1	1
Brighton	Fulham	H	1	0
Chelsea	Bournemouth	A	2	0
Chelsea	Bournemouth	D	2	0
Chelsea	Bournemouth	H	2	1

Table: Example of the expansion of count outcome for Poisson trick (First two games of the 2018/2019 season).

Parameter Estimation

- Two main methods employed:
  - Maximum Likelihood
  - Bayesian Method
- Notation:

Model	Parameter	MLE
Davidson	$(\boldsymbol{\alpha},\delta)$	$(\hat{\boldsymbol{\alpha}}, \hat{\delta_1})$
Davidson-Beaver	$(\boldsymbol{\alpha}, \delta, \theta)$	$(\hat{\boldsymbol{\alpha}}^*, \hat{\delta_2}, \hat{\theta})$

Data Set

- English Premier League - 2017/2018 season (football-data.co.uk, 2018)
- 20 teams; each team plays every other team twice
- 190 distinct comparisons (pairs)
- 380 matches in total

MLE - Davidson and Davidson-Beaver Models

- Estimates on log scale to enable useful comparisons (Turner *et al.*, 2018).
  'West Ham' set as reference: teams with positive estimates have higher worth and those with negative estimates have lower worth.
- χ^2 goodness-of-fit tests yielded p-values ¿ 0.60 proving fit of models to data.
- Estimates converged after 4 iterations with 1×10^{-6} tolerance.

Results

Team	$\hat{\alpha}_i$ (power:1/2)	α_i^* (power:1/2)	Team	$\hat{\alpha}_i$ (power:1/3)	$\hat{\alpha}_i{}^*$ (1/3)
Man City	3.816	4.012	Man City	3.400	3.573
Man United	2.116	2.241	Man United	2.046	2.162
Tottenham	1.907	2.021	Tottenham	1.821	1.926
Liverpool	1.907	2.021	Liverpool	1.712	1.812
Chelsea	1.427	1.515	Chelsea	1.448	1.534
Arsenal	0.988	1.051	Arsenal	1.091	1.157
Burnley	0.654	0.696	Burnley	0.636	0.675
Everton	0.327	0.348	Everton	0.378	0.401
Leicester	0.245	0.261	Leicester	0.273	0.289
Crystal Palace	0.082	0.087	Crystal Palace	0.111	0.118
Bournemouth	0.082	0.087	Newcastle	0.111	0.118
West Ham	0.000	0.00	Bournemouth	0.111	0.118
Newcastle	-0.000	-0.000	West Ham	0.000	0.000
Brighton	-0.083	-0.088	Watford	-0.057	-0.060
Watford	-0.166	-0.176	Brighton	-0.114	-0.121
Southampton	-0.249	-0.266	Huddersfield	-0.290	-0.308
Huddersfield	-0.334	-0.356	Southampton	-0.351	-0.373
Stoke	-0.506	-0.539	Swansea	-0.541	-0.574
West Brom	-0.595	-0.633	Stoke	-0.541	-0.574
Swansea	-0.595	-0.633	West Brom	-0.674	-0.715
$\hat{\delta}_1$ - Davidson draw	-0.122	-	$\hat{\delta}_1$	0.002	-
$\hat{\delta}_2$ - D-B draw	-	0.250	$\hat{\delta}_2$	-	0.380
$\hat{\theta}$ - home advantage	-	0.644	$\hat{\theta}$	-	0.644

Table: MLE estimates of the log strengths for models with and without home advantage effect.

Results (2)

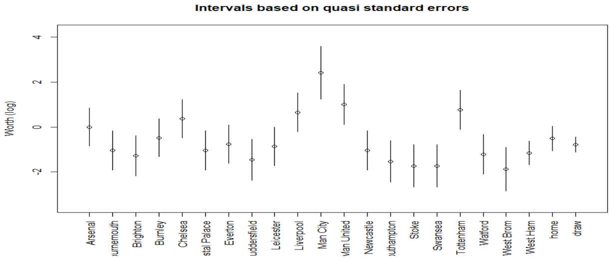


Figure: Log strengths for the Davidson-Beaver model with power 1/3 for 2017/2018.

- Intervals based on quasi-standard errors (Turner *et al.*, 2018).
- Non-overlapping intervals show significant differences in worth.

Outline

1. Classical Models and Results
2. **Bayesian Models and Results**
3. Model Comparisons

The general Bayesian model [1]

X_{ijr} is the outcome of the r^{th} match between team i and team j

$$\begin{aligned} \sigma &\sim \Gamma(2K, 2K/\hat{\sigma}^2) \\ \log(\boldsymbol{\alpha})|\sigma &\sim N(\mathbf{0}, \sigma^2\mathbf{I}) \\ \mathbf{p}_{ij} = \mathbf{p}_{ijr} &= f(\boldsymbol{\alpha}, \sigma, \ldots) \\ X_{ijr}|\boldsymbol{\alpha}, \sigma, \ldots &\sim \text{Multinomial}(1, \mathbf{p}_{ijr}) \end{aligned}$$

Where:

- ... denotes parameters specific to particular models
- Assume $\mathbf{p}_{ijr} = \mathbf{p}_{ij}$ for all r
- $\mathbf{p}_{ij} = (Pr(i \text{ beats } j), Pr(j \text{ beats } i), Pr(i \text{ draws } j))$
- $\hat{\sigma}$ is the estimated variance of $\log(\alpha_i)$
- K teams

[1] Phelan and Whelan (2017)

The general Bayesian model [1]

X_{ijr} is the outcome of the r^{th} match between team i and team j

$$\mathbf{p}_{ijr} = f(\boldsymbol{\alpha}, \sigma, \ldots)$$
$$X_{ijr} | \boldsymbol{\alpha}, \sigma, \ldots \sim \text{Multinomial}(1, \mathbf{p}_{ijr})$$

Where:

- ... denotes parameters specific to particular models

[1] Phelan and Whelan (2017)

The general Bayesian model [1]

X_{ijr} is the outcome of the r^{th} match between team i and team j

$$\mathbf{p}_{ij} = \mathbf{p}_{ijr} = f(\boldsymbol{\alpha}, \sigma, \ldots)$$
$$X_{ijr} | \boldsymbol{\alpha}, \sigma, \ldots \sim \text{Multinomial}(1, \mathbf{p}_{ijr})$$

Where:

- ... denotes parameters specific to particular models
- Assume $\mathbf{p}_{ijr} = \mathbf{p}_{ij}$ for all r

[1] Phelan and Whelan (2017)

The general Bayesian model [1]

X_{ijr} is the outcome of the r^{th} match between team i and team j

$$\mathbf{p}_{ij} = \mathbf{p}_{ijr} = f(\boldsymbol{\alpha}, \sigma, \dots)$$
$$X_{ijr}|\boldsymbol{\alpha}, \sigma, \dots \sim \text{Multinomial}(1, \mathbf{p}_{ijr})$$

Where:

- ... denotes parameters specific to particular models
- Assume $\mathbf{p}_{ijr} = \mathbf{p}_{ij}$ for all r
- $\mathbf{p}_{ij} = (Pr(i \text{ beats } j), Pr(j \text{ beats } i), Pr(i \text{ draws } j))$

[1] Phelan and Whelan (2017)

The general Bayesian model [1]

X_{ijr} is the outcome of the r^{th} match between team i and team j

$$\log(\boldsymbol{\alpha})|\sigma \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$
$$\mathbf{p}_{ij} = \mathbf{p}_{ijr} = f(\boldsymbol{\alpha}, \sigma, \ldots)$$
$$X_{ijr}|\boldsymbol{\alpha}, \sigma, \ldots \sim \text{Multinomial}(1, \mathbf{p}_{ijr})$$

Where:

- ... denotes parameters specific to particular models
- Assume $\mathbf{p}_{ijr} = \mathbf{p}_{ij}$ for all r
- $\mathbf{p}_{ij} = (Pr(i \text{ beats } j), Pr(j \text{ beats } i), Pr(i \text{ draws } j))$

[1] Phelan and Whelan (2017)

The general Bayesian model [1]

X_{ijr} is the outcome of the r^{th} match between team i and team j

$$
\begin{aligned}
\sigma &\sim \Gamma(2K, 2K/\hat{\sigma}^2) \\
\log(\boldsymbol{\alpha})|\sigma &\sim N(\mathbf{0}, \sigma^2\mathbf{I}) \\
\mathbf{p}_{ij} = \mathbf{p}_{ijr} &= f(\boldsymbol{\alpha}, \sigma, \ldots) \\
X_{ijr}|\boldsymbol{\alpha}, \sigma, \ldots &\sim \text{Multinomial}(1, \mathbf{p}_{ijr})
\end{aligned}
$$

Where:

- ... denotes parameters specific to particular models
- Assume $\mathbf{p}_{ijr} = \mathbf{p}_{ij}$ for all r
- $\mathbf{p}_{ij} = (Pr(i \text{ beats } j), Pr(j \text{ beats } i), Pr(i \text{ draws } j))$
- $\hat{\sigma}$ is the estimated variance of $\log(\alpha_i)$
- K teams

[1] Phelan and Whelan (2017)

Other priors

Models are differentiated by functions $f(\boldsymbol{\alpha}, \sigma, ...)$ for $\mathbf{p}_{ij}$

Table: Parameters of $f(\boldsymbol{\alpha}, \sigma, ...)$ and their priors

Model	**Draw**	**Home advantage**
Bradley-Terry	-	-
Davidson	$\delta \sim \text{Exp}(1)$	-
Davidson-Beaver	$\delta \sim \text{Exp}(1)$	$\theta \sim \text{Exp}(1)$
Rao-Kupper	$\bar{\gamma} \sim \text{Exp}(1)$	-
Rao-Kupper Mult	$\bar{\gamma} \sim \text{Exp}(1)$	$\theta \sim \text{Exp}(1)$

Note: $\bar{\gamma} = \gamma + 1$

Davidson Power Model

Question: How many points should be awarded for a win relative to a draw?
Answer: Let's look at one more model!
Davidson-Beaver:

$$P(i \text{ beats } j) = \frac{\theta\alpha_i}{\theta\alpha_i + \alpha_j + \delta \times (\alpha_i\alpha_j)^{1/2}}$$

$$P(j \text{ beats } i) = \frac{\alpha_j}{\theta\alpha_i + \alpha_j + \delta \times (\alpha_i\alpha_j)^{1/2}}$$

$$P(\text{draw}) = \frac{\delta \times (\alpha_i\alpha_j)^{1/2}}{\theta\alpha_i + \alpha_j + \delta \times (\alpha_i\alpha_j)^{1/2}}$$

Davidson Power Model

Question: How many points should be awarded for a win relative to a draw?
Answer: Let's look at one more model!
Davidson Power:

$$P(i \text{ beats } j) = \frac{\theta\alpha_i}{\theta\alpha_i + \alpha_j + \delta \times (\alpha_i\alpha_j)^\beta}$$
$$P(j \text{ beats } i) = \frac{\alpha_j}{\theta\alpha_i + \alpha_j + \delta \times (\alpha_i\alpha_j)^\beta}$$
$$P(\text{draw}) = \frac{\delta \times (\alpha_i\alpha_j)^\beta}{\theta\alpha_i + \alpha_j + \delta \times (\alpha_i\alpha_j)^\beta}$$

Fitting Bayesian Models

rStan package in R - wrapper for Stan (written in C++)

```
model {
  matrix[K, K] Sigma;

  // exp prior for delta (draws)
  delta ~ exponential(d);

  // exp prior for beta (power)
  beta ~ exponential(b);

  // exp prior for theta (home adv)
  theta ~ exponential(t);

  // gamma hyperprior for variance of lambdas
  sigma ~ gamma(2*K, 2*K/sigma_hat^2);
  Sigma = diag_matrix(rep_vector(sigma ^ 2, K));

  // multivariate norm prior for alpha (strength)
  lambda ~ multi_normal(rep_vector(0, K), Sigma);

  //loop over all the games
  for(r in 1:N){
    result[r] ~ categorical(outcome_probs[r]);
  }
}
```

Figure: Example rStan code for Davidson Power model

Evaluating Bayesian model fit

- Use the posterior distribution of parameters to generate replicate samples. This is the **posterior predictive distribution** (PPD) Gelman et al. (1996)
- **Classification rate** = proportion of the time the replicate samples correctly predict observed outcome
- **Other fit metrics** = use PPD as reference distribution for other test statistics derived from observed data (i.e. of draws, of wins by home team)

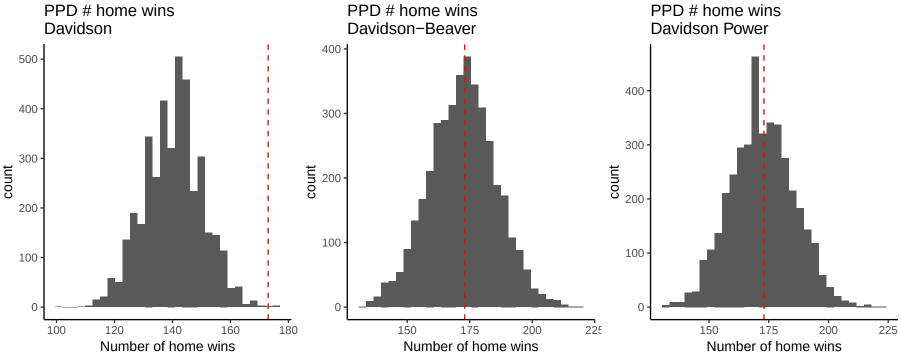
Evaluating Bayesian model fit

Classification rates:

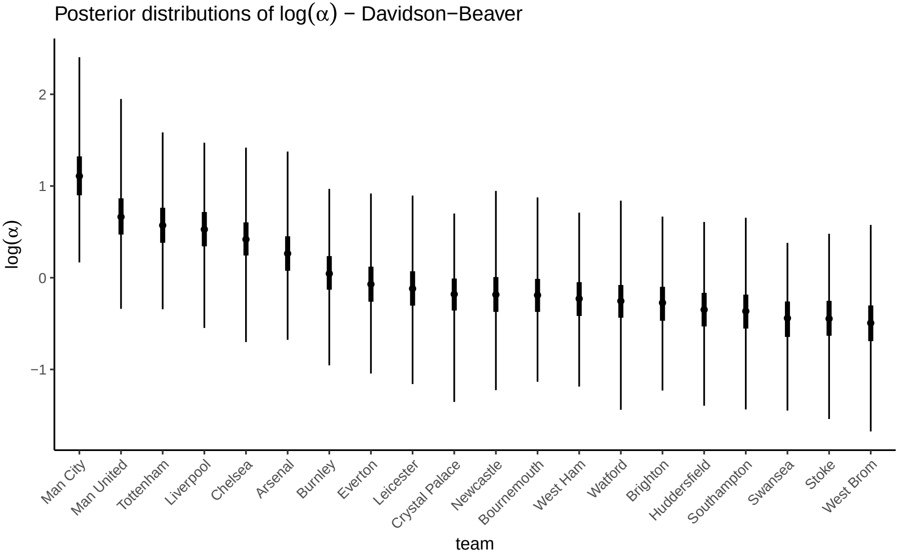
Model	PPD Classification Rate
Davidson	38.94 %
Davidson-Beaver	40.42 %
Davidson Power	40.63 %
Rao-Kupper	35.18 %
Rao-Kupper Mult	36.73 %

Evaluating Bayesian model fit

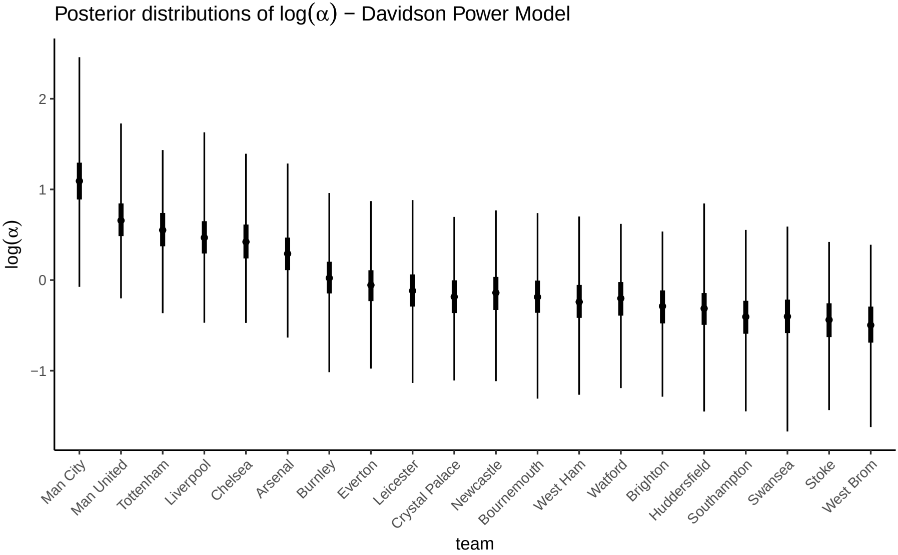
PPD home win distribution:



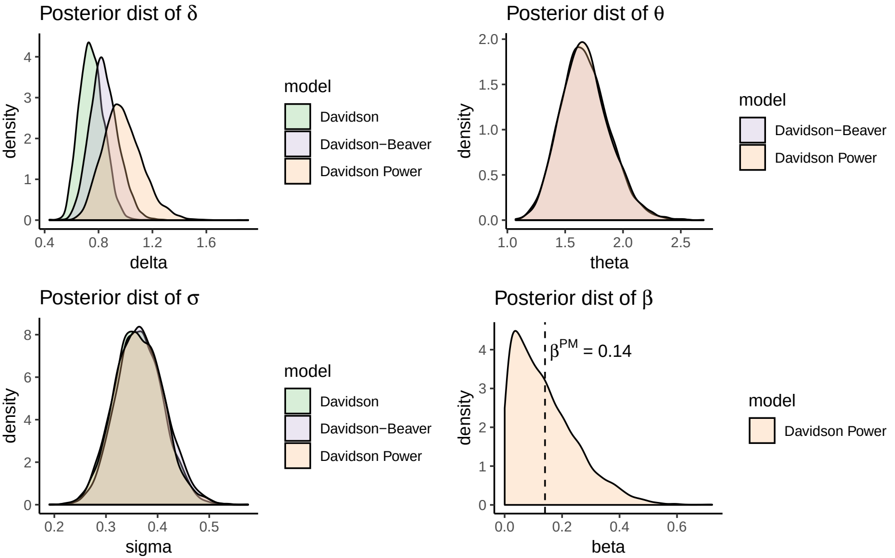
Bayesian Davidson Beaver Ranking - Season 2017/18



Bayesian Davidson Power Ranking - Season 2017/18



Parameter Comparison



Outline

1. Classical Models and Results
2. Bayesian Models and Results
3. **Model Comparisons**

Current Season Standings

Rank	Bayes Davidson (1/3)	Bayes Davidson (β)	Classical Davidson (1/3)
1	Man City	Man City	Man City
2	Liverpool	Chelsea	Liverpool
3	Chelsea	Liverpool	Tottenham
4	Tottenham	Arsenal	Chelsea
5	Arsenal	Tottenham	Arsenal
6	Watford	Everton	Man United
7	Man United	Bournemouth	Everton
8	Bournemouth	Man United	Wolves
9	Everton	Wolves	Watford
10	Leicester	Bournemouth	Bournemouth
11	Wolves	Leicester	West Ham
12	Brighton	Brighton	Leicester
13	West Ham	West Ham	Brighton
14	Burnley	Southampton	Newcastle
15	Newcastle	Burnley	Southampton
16	Crystal Palace	Newcastle	Burnley
17	Cardiff	Huddersfield	Crystal Palace
18	Southampton	Crystal Palace	Cardiff
19	Fulham	Cardiff	Huddersfield
20	Huddersfield	Fulham	Fulham

Questions

Thank you! Questions?

References

Andrew Gelman, Xiao-Li Meng, and Hal Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760, 1996.

Gabriel C Phelan and John T Whelan. Hierarchical bayesian bradley-terry for applications in major league baseball. *arXiv preprint arXiv:1712.05879*, 2017.