Having A Ball

Evaluating scoring streaks and game excitement using in-match trend estimation

Claus Thorn Ekstrøm and Andreas Kryger Jensen

March 24th 2021

Biostatistics, Department of Public Health
University of Copenhagen

Roadmap

1. Introduction
2. Latent dynamics GP Regression
3. The 2019–2020 NBA season
4. Conclusion

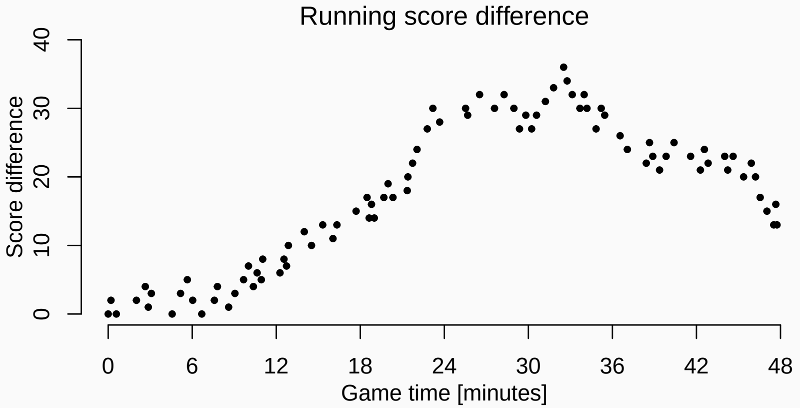
Introduction

How it all started

```
Dear sports-analytics-enthusiastic colleagues,

Today I want to make you aware of yet another call for a Special
Issue on Statistics in Sports in the AStA Advances in Statistical
Analysis journal (A Journal of the German Statistical Society)
which Dominik Liebl and I are currently hosting as guest editors:

https://www.springer.com/journal/10182/updates/17805238

We are looking forward to receiving high quality submission
from at least some of you until 31 October 2020. :-)

All the best, take care and stay safe,
Dominik & Andreas
```

Motivating example

Game development in the final match of the NBA 2019–2020 season between Los Angeles Lakers and Miami Heat at October 11, 2020. Positive values ⇒ LA Lakers are leading.



How exciting was this match?

Latent dynamics GP Regression

Gaussian Process regression with latent dynamics

The generative model of GP regression with data $\mathcal{D}_m = (D_m(t_{mi}), t_{mi}))_{i=1}^{n_m}$ is given hierarchically as

$$
\begin{aligned}
\faBasketballBall_m \mid \boldsymbol{\Psi}_m, \mathbf{t}_m &\sim H(\faBasketballBall_m \mid \boldsymbol{\Psi}_m) \\
d_m(t) \mid \faBasketballBall_m &\sim \mathcal{GP}(\mu_{\boldsymbol{\beta}_m}(t), C_{\boldsymbol{\theta}_m}(s,t)) \\
D_m(t_{mi}) \mid d_m(t_{mi}), t_{mi}, \faBasketballBall_m &\overset{iid}{\sim} N\left(d_m(t_{mi}), \sigma_m^2\right)
\end{aligned}
$$

Gaussian Process regression with latent dynamics

The generative model of GP regression with data $\mathcal{D}_m = (D_m(t_{mi}), t_{mi}))_{i=1}^{n_m}$ is given hierarchically as

$$
\begin{aligned}
\circledast_m \mid \mathbf{\Psi}_m, \mathbf{t}_m &\sim H(\circledast_m \mid \mathbf{\Psi}_m) \\
d_m(t) \mid \circledast_m &\sim \mathcal{GP}(\mu_{\boldsymbol{\beta}_m}(t), C_{\boldsymbol{\theta}_m}(s,t)) \\
D_m(t_{mi}) \mid d_m(t_{mi}), t_{mi}, \circledast_m &\overset{iid}{\sim} N\left(d_m(t_{mi}), \sigma_m^2\right)
\end{aligned}
$$

By linearity of differentiation and the properties of the Gaussian distribution we may augment the latent space as

$$
\begin{bmatrix} d_m(s) \\ d'_m(t) \\ d''_m(u) \end{bmatrix} \mid \circledast_m \sim \mathcal{GP}\left(\begin{bmatrix} \mu_{\boldsymbol{\beta}_m}(s) \\ \mu'_{\boldsymbol{\beta}_m}(t) \\ \mu''_{\boldsymbol{\beta}_m}(u) \end{bmatrix}, \begin{bmatrix} C_{\boldsymbol{\theta}_m}(s,s') & \partial_2 C_{\boldsymbol{\theta}_m}(s,t) & \partial_2^2 C_{\boldsymbol{\theta}_m}(s,u) \\ \partial_1 C_{\boldsymbol{\theta}_m}(t,s) & \partial_1\partial_2 C_{\boldsymbol{\theta}_m}(t,t') & \partial_1\partial_2^2 C_{\boldsymbol{\theta}_m}(t,u) \\ \partial_1^2 C_{\boldsymbol{\theta}_m}(u,s) & \partial_1^2\partial_2 C_{\boldsymbol{\theta}_m}(u,t) & \partial_1^2\partial_2^2 C_{\boldsymbol{\theta}_m}(u,u') \end{bmatrix} \right)
$$

(if the derivatives exist - no Ornstein-Uhlenbeck for you, Sir!)

Gaussian Process regression with latent dynamics

Inference about the latent processes is given by the posterior

$$\begin{bmatrix} d_m(s) \\ d'_m(t) \\ d''_m(u) \end{bmatrix} \mid \mathcal{D}_m, \circledast_m \sim \mathcal{GP}\left(\begin{bmatrix} \mu_{d_m}(s) \\ \mu_{d'_m}(t) \\ \mu_{d''_m}(u) \end{bmatrix}, \begin{bmatrix} \Sigma_{d_m}(s, s') & \Sigma_{d_m d'_m}(s, t) & \Sigma_{d_m d''_m}(s, u) \\ \Sigma_{d'_m d_m}(t, s) & \Sigma_{d'_m}(t, t') & \Sigma_{d'_m d''_m}(t, u) \\ \Sigma_{d''_m d_m}(u, s) & \Sigma_{d''_m d'_m}(u, t) & \Sigma_{d''_m}(u, u') \end{bmatrix} \right)$$

Gaussian Process regression with latent dynamics

Inference about the latent processes is given by the posterior

$$\begin{bmatrix} d_m(s) \\ d'_m(t) \\ d''_m(u) \end{bmatrix} \mid \mathcal{D}_m, \text{🏀}_m \sim \mathcal{GP}\left(\begin{bmatrix} \mu_{d_m}(s) \\ \mu_{d'_m}(t) \\ \mu_{d''_m}(u) \end{bmatrix}, \begin{bmatrix} \Sigma_{d_m}(s,s') & \Sigma_{d_m d'_m}(s,t) & \Sigma_{d_m d''_m}(s,u) \\ \Sigma_{d'_m d_m}(t,s) & \Sigma_{d'_m}(t,t') & \Sigma_{d'_m d''_m}(t,u) \\ \Sigma_{d''_m d_m}(u,s) & \Sigma_{d''_m d'_m}(u,t) & \Sigma_{d''_m}(u,u') \end{bmatrix} \right)$$

Where the components evaluated at any $\mathbf{t}^*$ are given by

$$\begin{aligned}
\mu_{d_m}(\mathbf{t}^*) &= \mu_{\boldsymbol{\beta}_m}(\mathbf{t}^*) + C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1}(\mathbf{D}_m - \mu_{\boldsymbol{\beta}_m}(\mathbf{t}_m)) \\
\mu_{d'_m}(\mathbf{t}^*) &= \mu'_{\boldsymbol{\beta}_m}(\mathbf{t}^*) + \partial_1 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1}(\mathbf{D}_m - \mu_{\boldsymbol{\beta}_m}(\mathbf{t}_m)) \\
\mu_{d''_m}(\mathbf{t}^*) &= \mu''_{\boldsymbol{\beta}_m}(\mathbf{t}^*) + \partial_1^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1}(\mathbf{D}_m - \mu_{\boldsymbol{\beta}_m}(\mathbf{t}_m)) \\
\Sigma_{d_m}(\mathbf{t}^*, \mathbf{t}^*) &= C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}^*) - C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1} C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}^*) \\
\Sigma_{d'_m}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1\partial_2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1} \partial_2 C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}^*) \\
\Sigma_{d''_m}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1^2\partial_2^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}^*) - \partial_1^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1} \partial_2^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}^*) \\
\Sigma_{d_m d'_m}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}^*) - C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1} \partial_2 C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}^*) \\
\Sigma_{d_m d''_m}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_2^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}^*) - C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1} \partial_2^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}^*) \\
\Sigma_{d'_m d''_m}(\mathbf{t}^*, \mathbf{t}^*) &= \partial_1\partial_2^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}^*, \mathbf{t}^*) - \partial_1 C_{\boldsymbol{\theta}}(\mathbf{t}^*, \mathbf{t}_m)\left(C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right)^{-1} \partial_2^2 C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}^*)
\end{aligned}$$

Trend Direction Index (TDI)

Through the posterior derivative we define the Trend Direction Index as

$$
\begin{aligned}
\mathrm{TDI}_m(t \mid \text{🏀}_m) &= P\left(d'_m(t) > 0 \mid \mathcal{D}_m, \text{🏀}_m\right) \\
&= \frac{1}{2} + \frac{1}{2}\,\mathrm{Erf}\left(\frac{\mu_{d'_m}(t)}{2^{1/2}\Sigma_{d'_m}(t,t)^{1/2}}\right)
\end{aligned}
$$

The probability that the score difference is increasing at time t. Switching reference team leads to $1 - \mathrm{TDI}_m(t \mid \text{🏀}_m)$.

Excitement Trend Index

We further derfine The Excitement Trend Index as an application of Rice's formula (1936)

$$
\begin{aligned}
\mathrm{ETI}_m \mid \text{🏀}_m &= \mathbb{E}\left[\#\left\{t \in \mathcal{I}_m : d'_m(t) = 0\right\} \mid \mathcal{D}_m, \text{🏀}_m\right] \\
&= \int_{\mathcal{I}_m} \int_{-\infty}^{\infty} |v| \, f_{d'_m(t), d''_m(t)}(0, v \mid \mathcal{D}_m, \text{🏀}_m) \mathrm{d}v \mathrm{d}t \\
&= \int_{\mathcal{I}_m} d\mathrm{ETI}_m(t \mid \text{🏀}_m) \mathrm{d}t
\end{aligned}
$$

The expected number of times that the score difference changes monotonicity. Invariate to switching reference team.

Excitement Trend Index

The local Excitement Trend Index has the following explicit expression:

$$d\mathrm{ETI}_m(t \mid \text{🏀}_m) = \lambda_m(t)\phi\left(\frac{\mu_{d'_m}(t)}{\Sigma_{d'_m}(t,t)^{1/2}}\right)\left(2\phi\left(\zeta_m(t)\right) + \zeta_m(t)\,\mathrm{Erf}\left(\frac{\zeta_m(t)}{2^{1/2}}\right)\right)$$

with

$$\lambda_m(t) = \frac{\Sigma_{d''_m}(t,t)^{1/2}}{\Sigma_{d'_m}(t,t)^{1/2}}\left(1-\omega_m(t)^2\right)^{1/2}, \quad \omega_m(t) = \frac{\Sigma_{d'_m d''_m}(t,t)}{\Sigma_{d'_m}(t,t)^{1/2}\Sigma_{d''_m}(t,t)^{1/2}}$$

$$\zeta_m(t) = \frac{\mu_{d'_m}(t)\Sigma_{d''_m}(t,t)^{1/2}\omega_m(t)\Sigma_{d'_m}(t,t)^{-1/2} - \mu_{d''_m}(t)}{\Sigma_{d''_m}(t,t)^{1/2}\left(1-\omega_m(t)^2\right)^{1/2}}$$

Estimation

$$\mu_{\boldsymbol{\beta}_m}(t) = \beta_m, \quad C_{\boldsymbol{\theta}_m}(s,t) = \alpha_m^2 \exp\left(-\frac{(s-t)^2}{2\rho_m^2}\right)$$

The hyper-parameters are given the following prior distribution

$$H(\boldsymbol{\theta}_m \mid \boldsymbol{\Psi}_m) = H(\beta_m \mid \Psi_{\beta_m})H(\alpha_m \mid \Psi_{\alpha_m})H(\rho_m \mid \Psi_{\rho_m})H(\sigma_m \mid \Psi_{\sigma_m})$$

with

$$\beta_m \sim T_4\left(\widehat{\beta_m^{\mathrm{ML}}}, 5\right),\ \alpha_m \sim T_4^+\left(\widehat{\alpha_m^{\mathrm{ML}}}, 5\right),\ \rho_m \sim T_4^+\left(\widehat{\rho_m^{\mathrm{ML}}}, 5\right),\ \sigma_m \sim T_4^+\left(\widehat{\sigma_m^{\mathrm{ML}}}, 5\right)$$

Super-script ML indicates the marginal maximum-likelihood estimate (empirical Bayes).

Empirical bayes

To find the empirical Bayes estimate of the hyper-parameters we maximize the marginal log-likelihood function. This can be written as

$$\begin{aligned}\log L(\text{🏀}_m \mid \mathcal{D}_m) \propto& -\frac{1}{2}\log\left|C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right| \\ &- \frac{1}{2}(\mathbf{D}_m - \mu_{\boldsymbol{\beta}_m}(\mathbf{t}_m))^T \left[C_{\boldsymbol{\theta}_m}(\mathbf{t}_m, \mathbf{t}_m) + \sigma_m^2 I\right]^{-1}(\mathbf{D}_m - \mu_{\boldsymbol{\beta}_m}(\mathbf{t}_m))\end{aligned}$$

The solution $\widehat{\text{🏀}_m^{\mathrm{ML}}} = \arg\sup_{\text{🏀}} \log L(\text{🏀} \mid \mathcal{D}_m)$ can be obtained by numerical optimization.

Our model is implemented in Stan and for each match we ran 4 chains for 50,000 iterations with half of them for warm-up. Output: 2292 gigabyes of posterior data.

The 2019–2020 NBA season

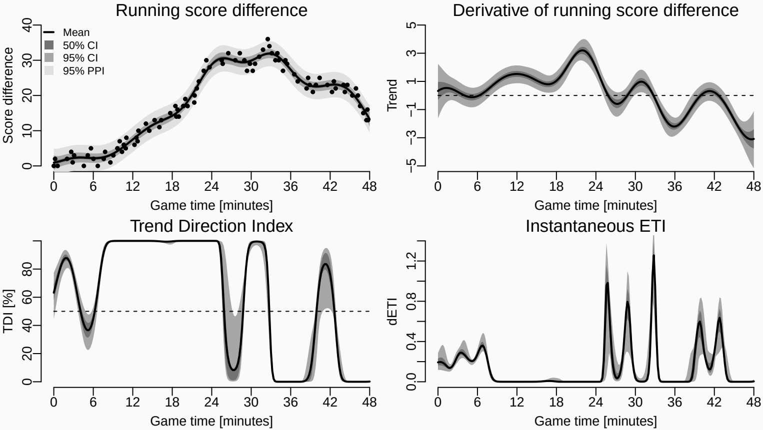
The 2019–2020 NBA season

We applied our method for all 1143 matches from the 2019–2020 NBA basketball season.

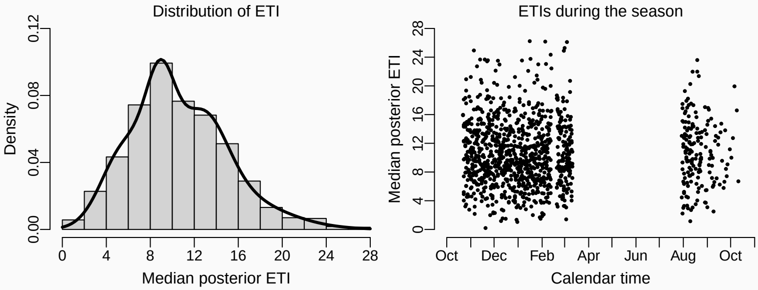
We disregarded overtime.

Final match between LA Lakers and Miami Heat

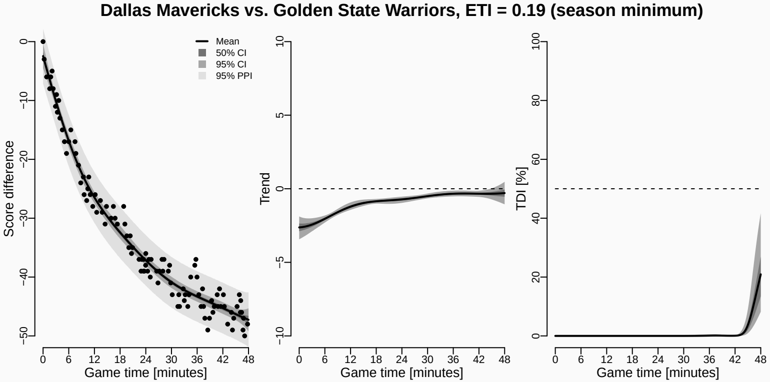
Analysis of the motivating example.



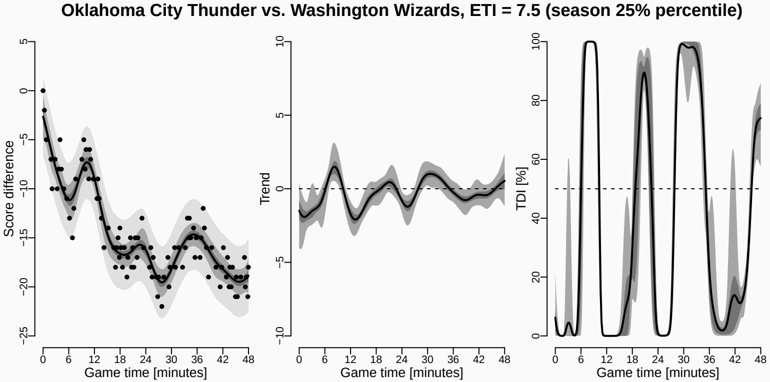
Distribution of the 1143 median posterior ETIs



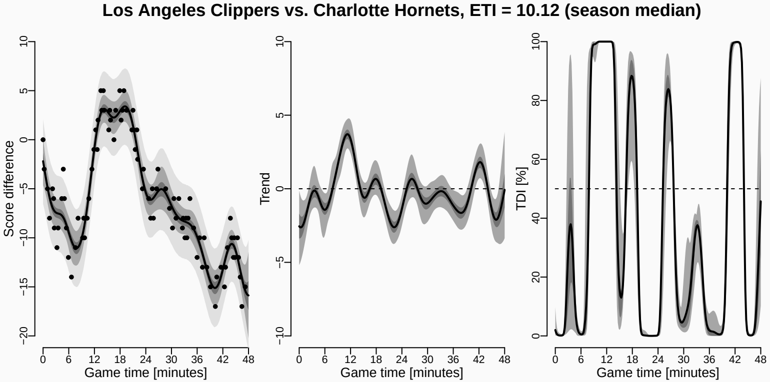
Season minimum ETI



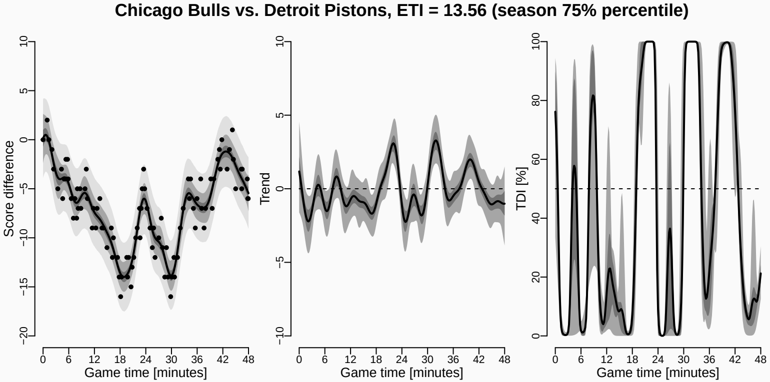
Season 25% percentile ETI



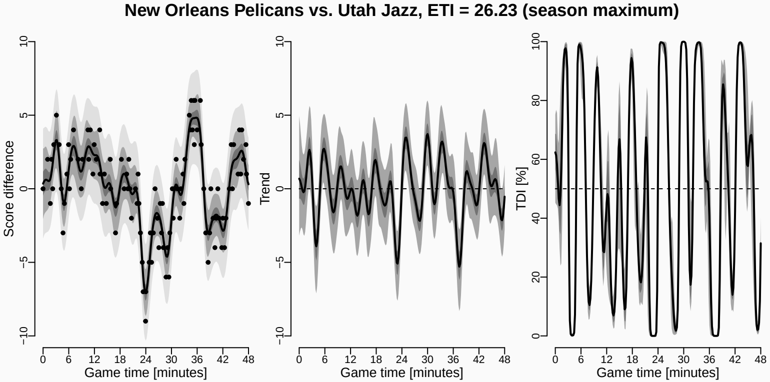
Season median ETI



Season 75% percentile ETI



Season maximum ETI



An algorithm for collapsing a factor into sub-groups

If I were to go see an exciting match — who should I watch?

An algorithm for collapsing a factor into sub-groups

If I were to go see an exciting match — who should I watch?

Cluster teams based on their median posterior ETI.

Compute the $\text{RMSEP}^{C=c}_{\text{LOO-CV}}$ where c is the number of subgroups considering all possible partitions of the 30 teams. Essentially a bunch of one-way ANOVAs

$\text{RMSEP}^{C=2}_{\text{LOO-CV}} = 4.493$

$\text{RMSEP}^{C=3}_{\text{LOO-CV}} = 4.49$

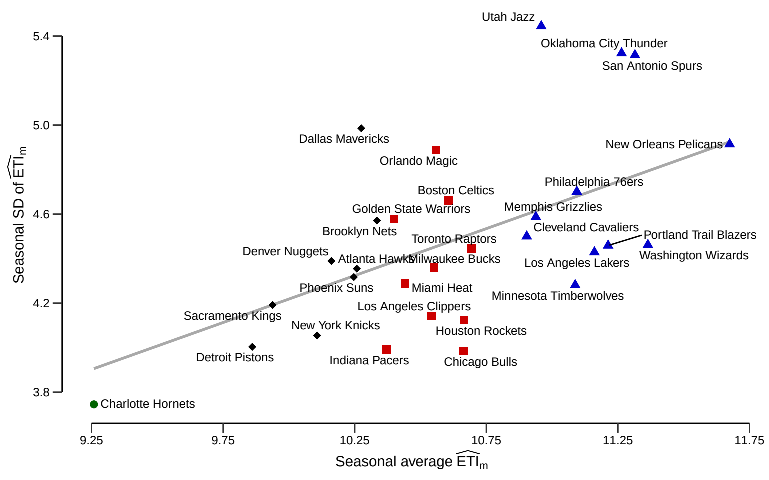
$\text{RMSEP}^{C=4}_{\text{LOO-CV}} = 4.489$

$\text{RMSEP}^{C=5+}_{\text{LOO-CV}} = 4.489$

Clustering results

	Average	SD	2.5%	50%	97.5%	Group
New Orleans Pelicans	11.67	4.91	3.10	11.59	23.54	A
Washington Wizards	11.36	4.46	2.98	11.63	18.28	A
San Antonio Spurs	11.31	5.31	3.52	9.88	23.44	A
Oklahoma City Thunder	11.26	5.32	1.68	11.55	23.74	A
Portland Trail Blazers	11.21	4.46	3.31	10.82	18.48	A
Los Angeles Lakers	11.16	4.43	3.61	10.52	19.79	A
Philadelphia 76ers	11.09	4.70	2.73	10.28	21.53	A
Minnesota Timberwolves	11.09	4.28	4.01	10.69	19.46	A
Utah Jazz	10.96	5.44	3.09	9.76	23.54	A
Memphis Grizzlies	10.94	4.59	3.04	10.36	20.84	A
Cleveland Cavaliers	10.90	4.50	4.93	10.39	22.36	A
Toronto Raptors	10.69	4.44	3.48	10.33	21.79	B
Houston Rockets	10.67	4.12	3.52	10.74	19.17	B
Chicago Bulls	10.66	3.98	2.95	10.65	17.49	B
Boston Celtics	10.61	4.66	2.81	10.19	20.92	B
Orlando Magic	10.56	4.89	2.72	10.41	21.01	B
Milwaukee Bucks	10.55	4.36	2.77	9.59	18.13	B
Los Angeles Clippers	10.54	4.14	4.05	10.72	20.56	B
Miami Heat	10.44	4.29	3.28	9.91	18.54	B
Golden State Warriors	10.40	4.58	3.37	10.24	18.72	B
Indiana Pacers	10.37	3.99	3.31	10.22	18.54	B
Brooklyn Nets	10.33	4.57	3.07	10.06	18.89	C
Dallas Mavericks	10.27	4.99	3.00	9.04	22.09	C
Atlanta Hawks	10.26	4.35	2.50	9.75	18.31	C
Phoenix Suns	10.25	4.32	3.78	9.29	19.81	C
Denver Nuggets	10.16	4.39	3.42	9.59	19.61	C
New York Knicks	10.11	4.05	2.84	9.87	18.47	C
Sacramento Kings	9.94	4.19	3.66	9.30	18.93	C
Detroit Pistons	9.86	4.00	3.73	9.44	17.08	C
Charlotte Hornets	9.26	3.74	1.96	9.22	15.95	D

Team specific seasonal SD vs. average ETIs



Conclusion

Conclusion

- The trender enables in-game and postgame evaluations about the underlying trends in scoring patterns.
- ETI is the expected number of monotonicity changes — which team appears stronger. High numbers good.

Future ideas:

- Use a weighted ETI to weigh the ETI higher towards the end of the game — or lower if one team is far ahead.

Also: have cemented our position as sports-analytics-enthusiasts.