Bayesian Sports Analytics

Tom Jeon & Steele Valenzuela

Motivation

- **Sports analytics:** Who's a better team, athlete, coach? Who'll win?
- **Our approach:** Model expected score differential, then map it back to probability of winning.
    - Consider very close games or blowouts; the win/loss outcome provides essentially zero information. There's a lot of information in the score differential that's thrown away if you just look at win/loss.
- **Original plan:** Create a Bayesian AR(1) model for expected score differential. Does the score differential at t = 10 minutes affect score differential at t = 20 minutes? Are teams motivated/demotivated by the present score differential in game?
- **Problem:** No data available in minute time lags. What did we do?

Overview

- Data set: NBA seasons from 2003-2004 to 2014-2015
- Specify Bayesian AR(1) model for observed data of 2014-2015
- Spline model
- Bayesian AR(1) of spline residuals
- Bayesian AR(1) of compiled seasons for the Boston Celtics
- Conclusion
- Questions

Dataset: NBA seasons from 2003 to 2016

How did we collect the data?

- rvest library was extremely helpful
- Scraped season schedule and results from sports-reference.com
- Merged team attributes and ratings

Cleaning

- dplyr & tidyr libraries
- Cleaning involved:
  - renaming variables and deleting columns
  - Switching from wide to long format
  - Filtering data for team additions, name changes, etc.
- Created ranking, an all encompassing predictor variable for score differential.

Before

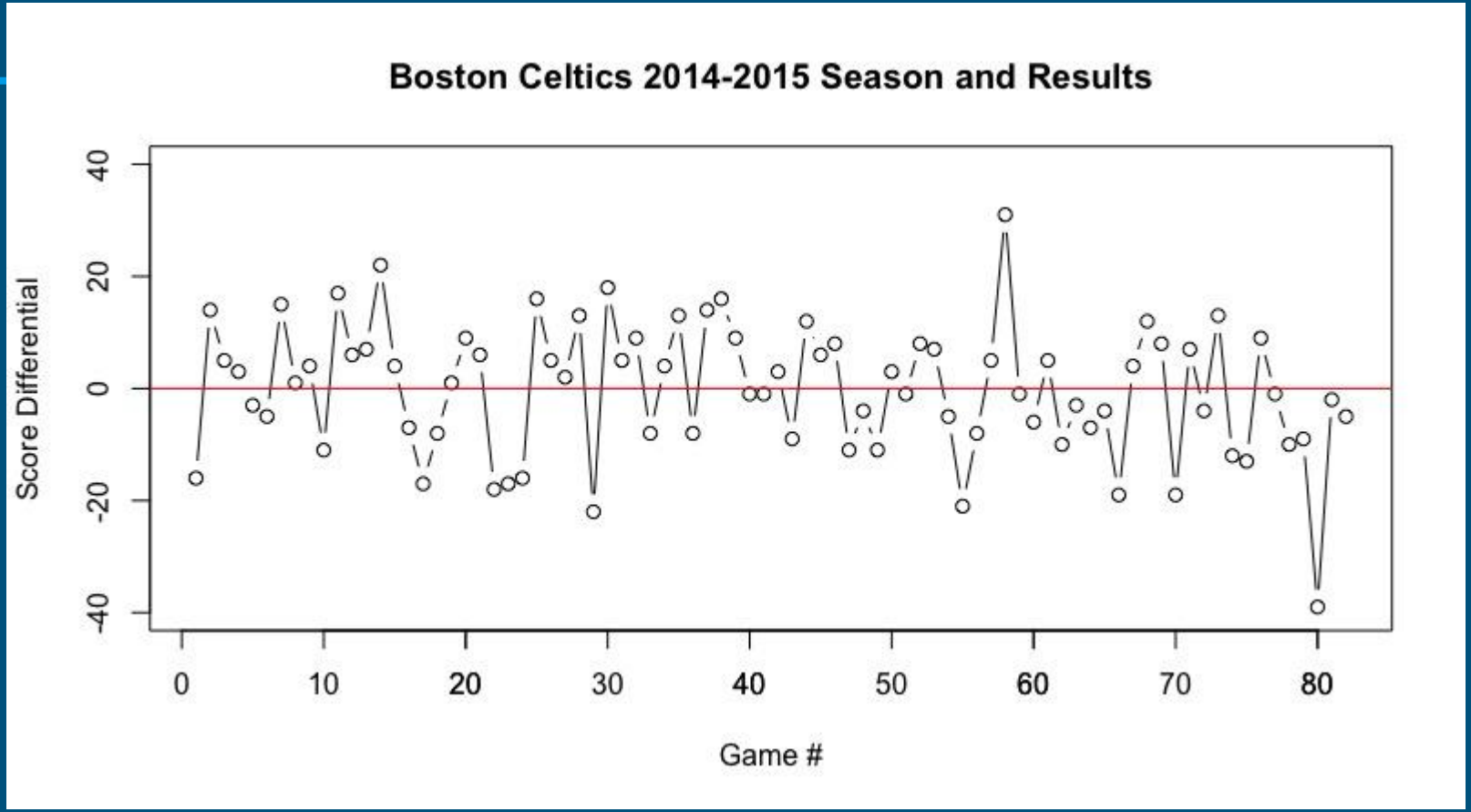
Regular Season Glossary · SHARE · Embed · CSV · Export · PRE · LINK · ?

Date	Start (ET)		Visitor/Neutral	PTS	Home/Neutral	PTS
			October			
Tue, Oct 27, 2015	8:00 pm	Box Score	Detroit Pistons	106	Atlanta Hawks	94
Tue, Oct 27, 2015	8:00 pm	Box Score	Cleveland Cavaliers	95	Chicago Bulls	97
Tue, Oct 27, 2015	10:30 pm	Box Score	New Orleans Pelicans	95	Golden State Warriors	111
Wed, Oct 28, 2015	7:30 pm	Box Score	Philadelphia 76ers	95	Boston Celtics	112
Wed, Oct 28, 2015	7:30 pm	Box Score	Chicago Bulls	115	Brooklyn Nets	100
Wed, Oct 28, 2015	7:30 pm	Box Score	Utah Jazz	87	Detroit Pistons	92

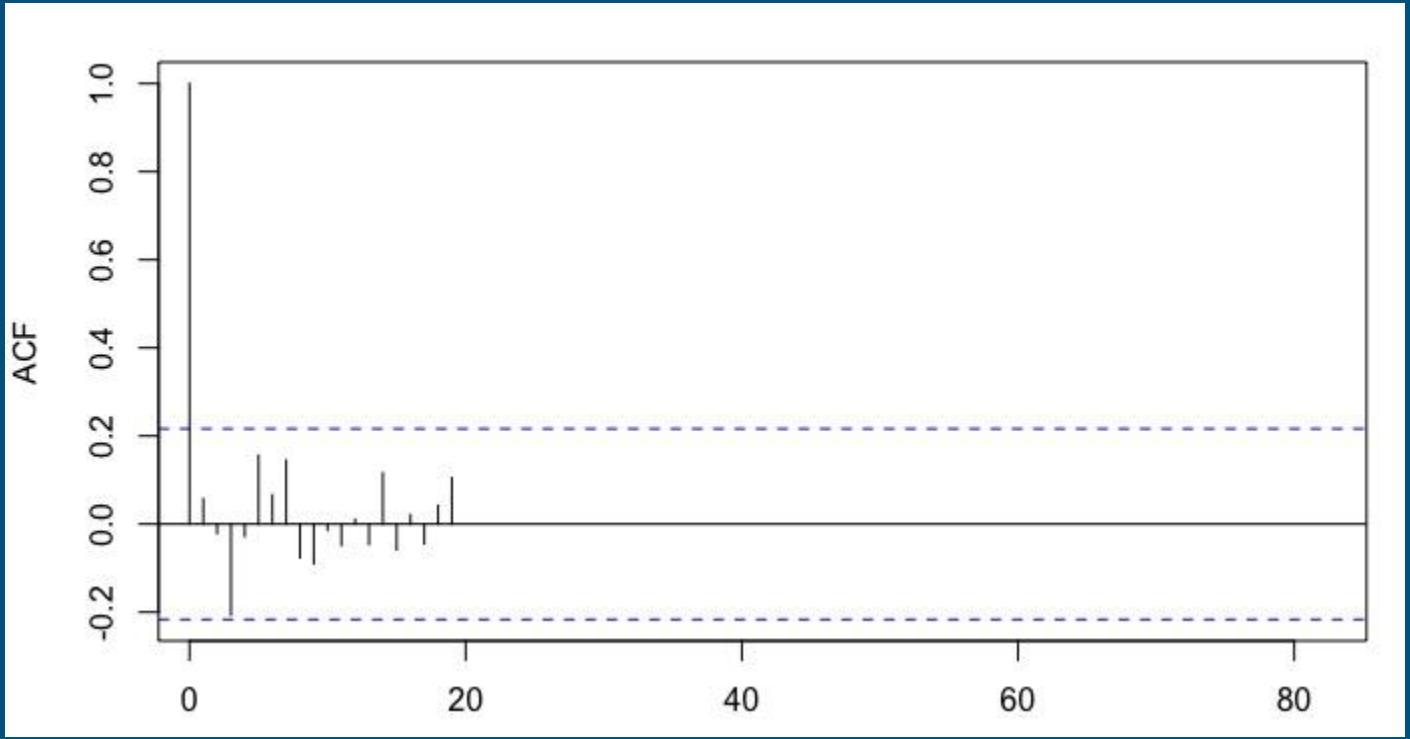
After

	Date	Score Differential	Opponent	Rank
1	Fri, Oct 31, 2003	2	Memphis Grizzlies	22
2	Sat, Nov 1, 2003	-7	New Orleans Hornets	14
3	Wed, Nov 5, 2003	-8	Detroit Pistons	5
4	Tue, Nov 11, 2003	2	Indiana Pacers	8
5	Mon, Nov 17, 2003	-3	New York Knicks	21
6	Fri, Nov 21, 2003	-2	Philadelphia 76ers	12

Let’s focus on the 2014-2015 Season



Well...



Let’s fit a Bayesian AR(1) model on this data

- First, our proposed model:

$$y_t \sim AR(1)$$
$$\rho \sim U(-1, 1)$$
$$\sigma \sim U(0, 15)$$

- Next, we used a stationary distribution for y_1:

$$y_1 \sim N(0, \sigma^2/(1-\rho^2))$$

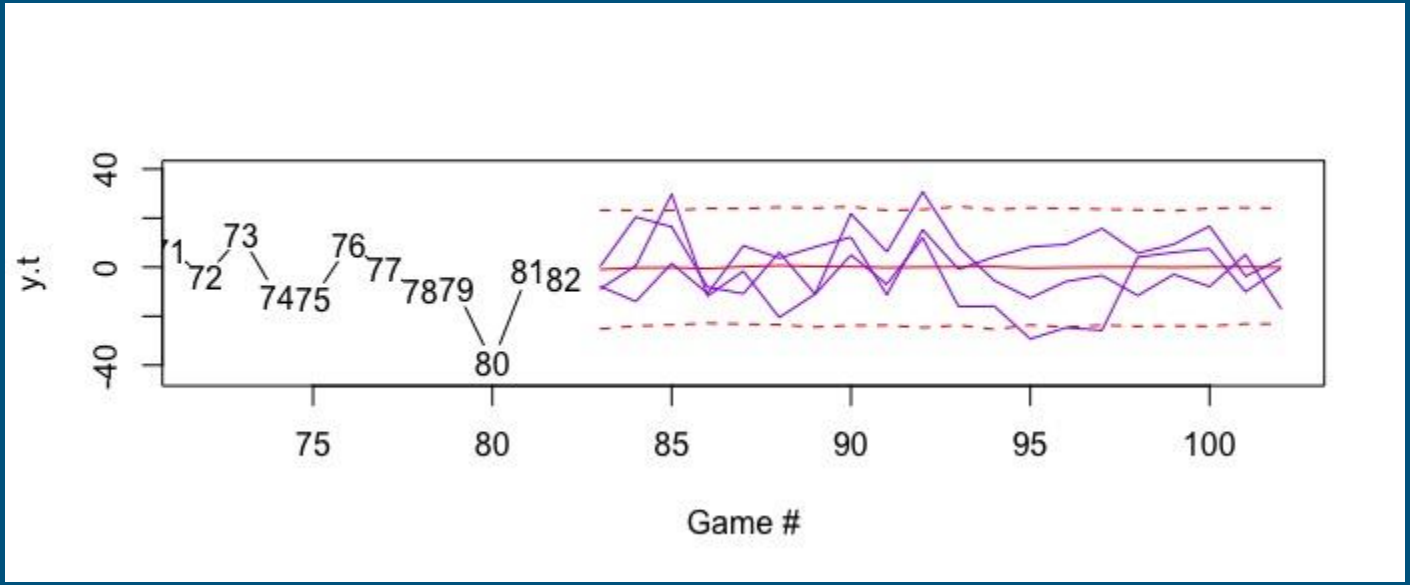
Model Fit

Here's how the model fared:

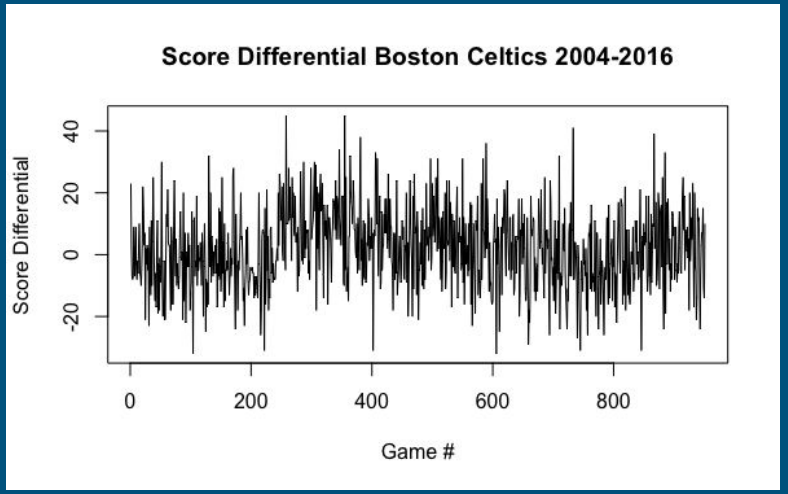
- Convergence & trace plots? ✔
- True parameters & posteriors? ✔

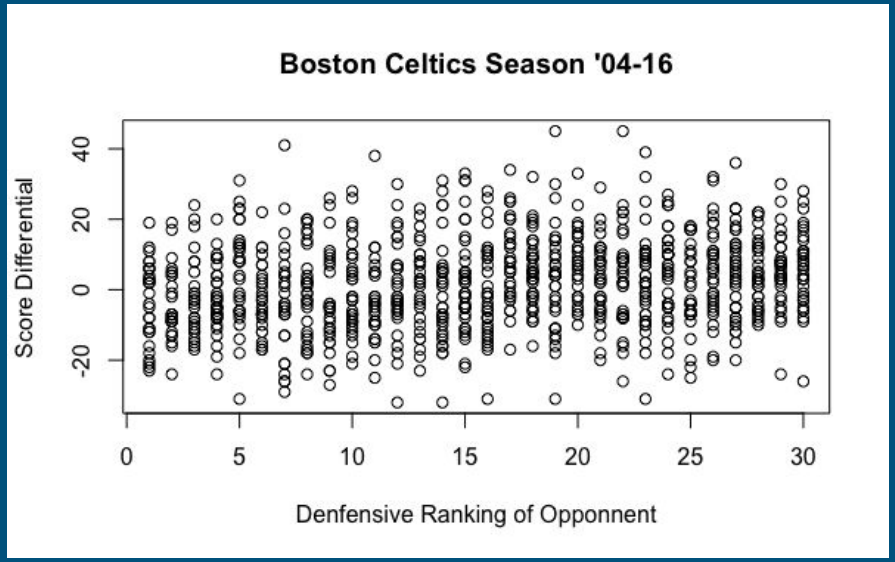
	true.param	mean	sd	Lower.Bound	Upper.Bound	tp.Lower.Bound	tp.Upper.Bound
sigma	11.632	11.895	0.962	10.181	13.960	NA	NA
rho	0.058	0.059	0.114	-0.173	0.282	-0.15956	0.27556

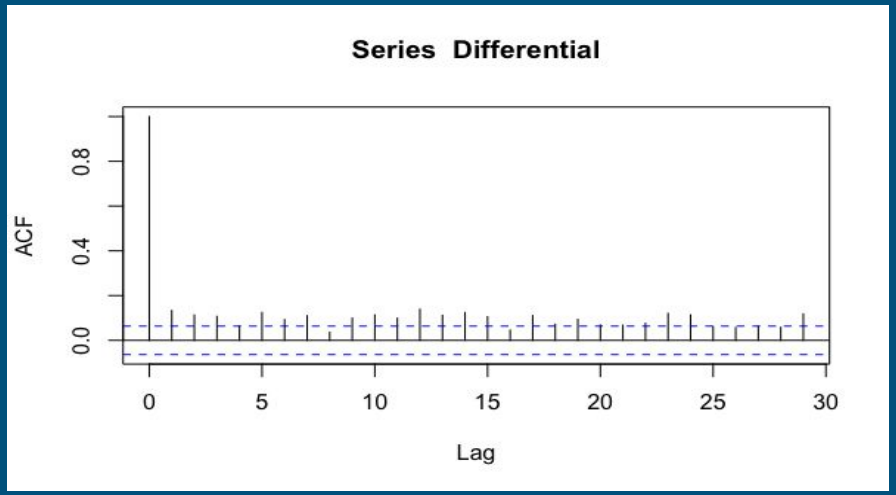
Forecast (who cares?)



Exploratory Analysis







Analytics: Who’s a better head coach?

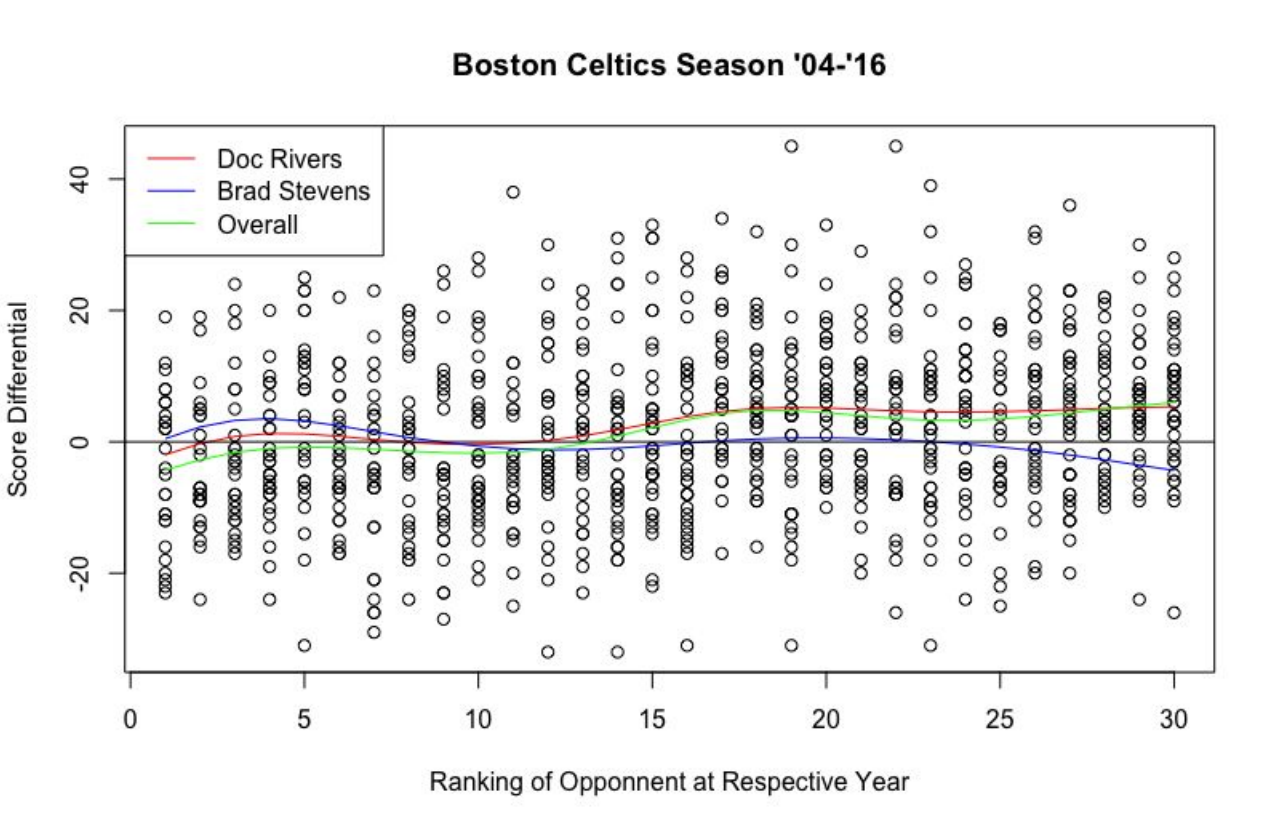
Let’s go into the specifics

- We used a splines model for when Doc Rivers was head coach and for the current head coach, Brad Stevens.
- Fit separate cubic splines models with knot intervals set equal to 5.
- Model specification:

$$
\begin{aligned}
y_i &\sim N(\mu_i, \sigma_y^2), \\
\mu_i &= \sum_{k=1}^{K} b_k(x_i)\alpha_k,
\end{aligned}
$$

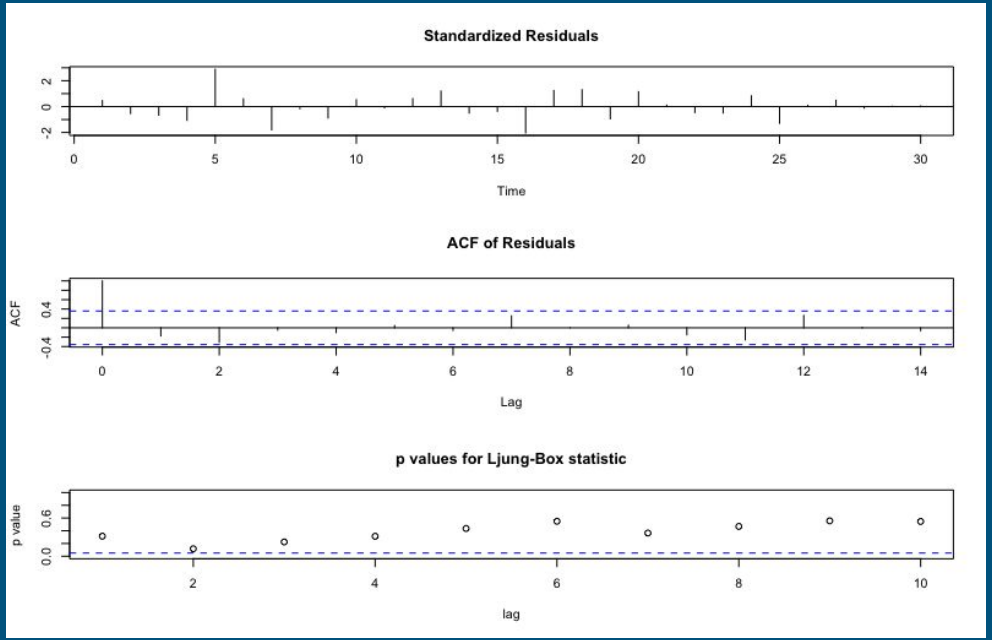
$$
\begin{aligned}
\alpha_k &\sim N(0, 100), \text{ indep. prior for each } k, \\
\sigma_y &\sim U(0, 3).
\end{aligned}
$$

Analytics: Who's a better coach?



Bayesian AR(1) model on spline residuals

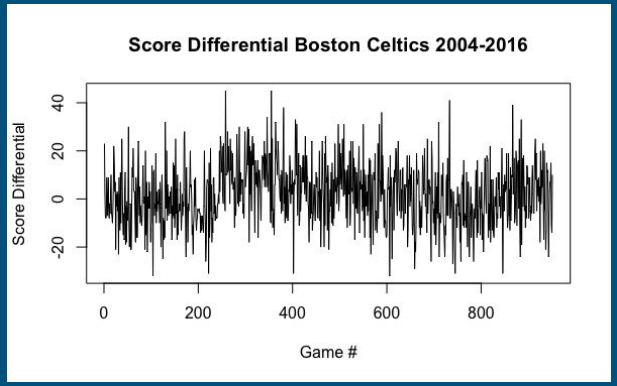
- To capture a bit of what we didn't capture with the spline model
- Diagnostics:



Bayesian AR(1) model on compiled seasons

- Again, our proposed model:

$$y_t \sim AR(1)$$
$$\rho \sim U(-1, 1)$$
$$\sigma \sim U(0, 15)$$



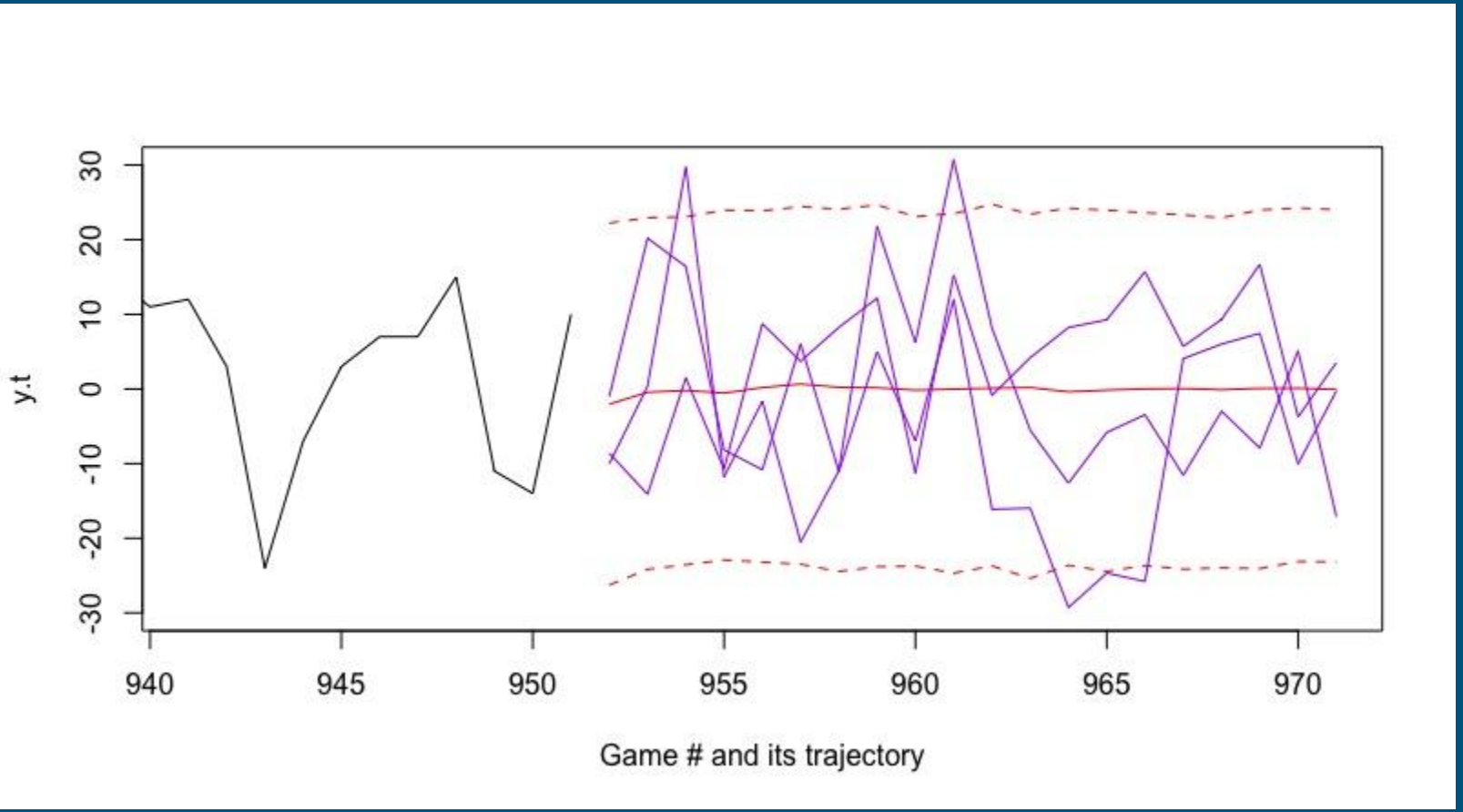
- And again, we used a stationary distribution for y_1:

$$y_1 \sim N(0, \sigma^2/(1-\rho^2))$$

Model fit on compiled season

	true.param	mean	sd	Lower.Bound	Upper.Bound	tp.Lower.Bound	tp.Upper.Bound
sigma	12.566	12.677	0.299	12.117	13.273	NA	NA
rho	0.134	0.150	0.032	0.088	0.213	0.07128	0.19672

Forecast of compiled seasons



Questions?