Are hot streaks in baseball pitchers real? Mhmm.

Abstract

In the past, "Hot streaks" in sports have been notoriously difficult to demonstrate despite the fact that many players across many different sports are convinced that they exist. Here we consider looking for streakiness in starting pitchers in major league baseball by examining a statistic that the pitcher has almost complete control over: fastball velocity. We fit a mixed hidden Markov model (HMM) in a Bayesian framework using the sequences of fastball velocities generated by starting pitchers. This model allows for parameters of the emission distributions and the transition matrices to vary by player while simultaneously allowing for pooling information across players. We begin by fitting a two state model and find evidence that there is approximately a 2-3 MPH difference between the states, on average. Further, we study the predictive effects of our model and find that "hot" pitchers outperform "cold" pitchers in future starts in such statistical categories such as strikeouts and ERA. We also find evidence that when a pitcher is "hot" based on their fastball that other pitches are also improved (e.g. curveball).

1 Introduction

1.1 The "Hot Hand"

There is a rich collection of literature studying the concept of "streakiness" or "hot hand" effect in sports and many classic research papers looking into the topic have concluded that this effect does not actually exist. The seminal work in this area is Gilovich and Tversky (1985), which studied the "hot hand" in basketball using shot data from the Philadelphia 76ers and free throws from the Boston Celtics, both of the National Basketball Association (NBA), along with free throw data collected under experimental conditions from Cornell's men's and women's varsity basketball teams. Based on their analysis they conclude that the hot hand in shooting does not exist, even referring to it as a "cognitive illusion" (Gilovich and Tversky (1985), Page 313). Additional work in this line of research supporting the idea that the hot hand is merely an illusion can be found in Huizinga and Weil (2009), Tversky and Gilovich (1989), Koehler and Conley (2003). More recently, however, the orthodoxy that the "hot hand" is not real is beginning to challenged as access to newer and larger data becomes available to researchers. For example, Arkes (2010) uses data from every free throw occurring in the 2005-06 season of the National Basketball Association (NBA), and they find a 2-3% increase in the probability of making the second free throw given that the first free three is made. Arkes (2013), continuing to focus on free throws, performs a simulation study to demonstrate that the primary methods (e.g. runs test and tests for stationarity) for detecting the hot hand in past studies (such as Huizinga and Weil (2009), Gilovich and Tversky (1985), Tversky and Gilovich (1989), Koehler and Conley (2003)) may not be adequate for detecting the hot hand. For instance, Arkes (2013) estimates that if the frequency of the hot hand is 20%, the true hot hand effect was 10 percentage points and the overall probability of making a free throw is 75% (i.e. probability of making a shot when hot is 83%; probability of making a shot when cold is 73%), there is only a 5.2% chance that the hot hand will be detected with a sample size of 2,160. When the sample size is increased to 28,800, there is still only a 33.0% chance of detecting the hot hand under these parameter conditions.

Moving beyond only free throws in basketball, Bocskocsky and Carolyn (2014) looked at approximately 83,000 field goal attempts from the 2012-2013

NBA season. After conditioning on the difficulty of the shot, they estimate the effect of the hot hand to be between 1.2 and 2.4 percentage points.

So we have longitudinal data from multiple pitchers. And we are hypothesizing that this data is generated from one of two hidden states. We could view this data as arising from a mixture model however, that would ignore the time dependence of the data. HMM allow for modeling data arising from a mixture distribution along with serial dependence in the data. A single HMM is therefore well suited for modeling each time series of each pitcher performance individually. However, our goal is to pool all these time series together to "borrow information" across pitchers to estimate league wide parameters for hot and cold.

1.2 Mixture Models, Hidden Markov Models, and Mixed Hidden Markov Models

Mixture models (CITE) fit data that are assumed to be generated from multiple processes, which is useful in the current application, as it is hypothesized that the data (i.e. adjusted fastball velocities) are being generated from two distinct states (i.e. "hot" and "cold"). However, mixture models assume independence between observations and do not account for serially correlated data.

Humphreys (1998) and Seltman (2002) both present work that deals with multiple sequences at once. Later, in Altman (2007) the methods described in the aforementioned articles are formalized and the term mixed hid-

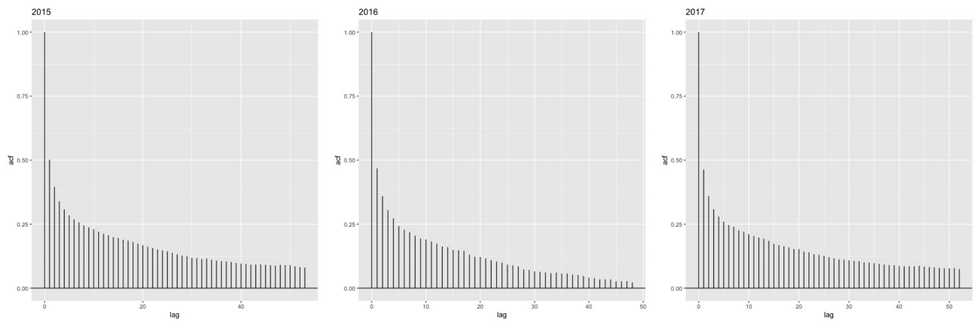


Figure 1: Autocorrelation plot for 2015, 2016, and 2017 for adjusted fastball velocities (i.e. residuals from the linear model.

den Markov models (MHMM) is applied to describe models of this type. Further work applying this method can be found in Shirley, Small, Lynch, Maisto, and Oslin (2010), which presents an application of the MHMM framework in the context of a clinical trial. An overview of MHMM can be found in Maruotti (2011).

Altman (2007) Meeden and Vardeman (1997)

2 Data

For pitchers we are looking at fastball velocities. We are using data from the DATA BASE NAME. We looked at starting pitchers from the major league baseball (MLB) during the regular season and we look at the 2015, and 2016, and 2017 seasons separately. In this analysis we included all pitches where the pitch type is listed as a fastball or pitch type is listed as a cutter and is classified as hard.

This data set contains 308 unique pitchers who threw a total of 253,462 pitches. After removing observation with missing values, 253,244 observations remain. We then filter to include only pitchers who threw more than 800 pitches during the 2016 season. This leaves us with 136 pitchers and 199,975 pitches.

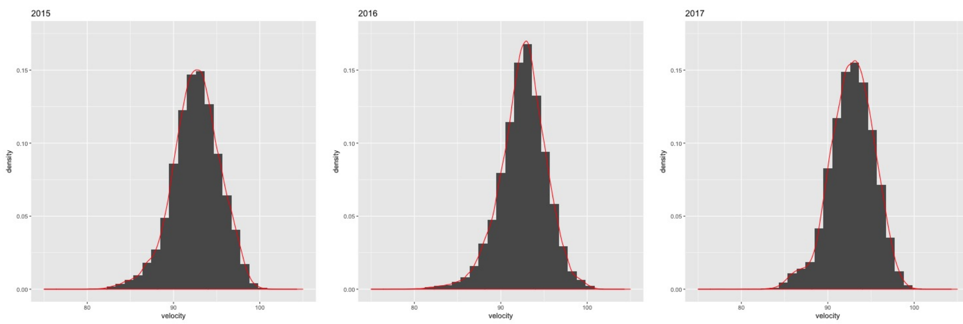


Figure 2: Histogram of velocities for 2015, 2016, and 2017

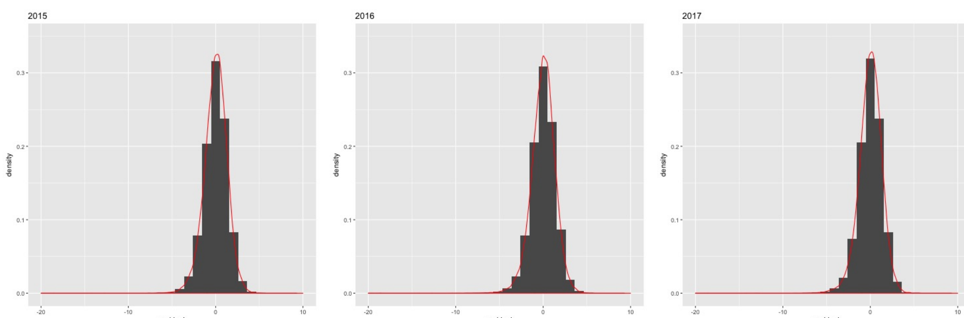


Figure 3: Histogram of residuals for 2015, 2016, and 2017

2015 data: 263186 observations After cleaning: 212190.

2016 Data: 253462 After cleaning 199975

2017 Data: 244057 After cleaning: 190529

3 Model

3.1 Controlling external effect

The overall goal of this model is to measure for a pitcher performs relative to their own average performance. That is, this model seeks to find strings of pitches where a pitcher is better than their average (i.e. “hot”) or worse than their average (i.e. “cold”). Therefore, a model to control for effects that are external to the pitcher is fit first prior to the HMM. This model is a linear mixed effects model with fixed effects for pitch type, pitch count, an indicator for baserunners, elevation, and temperature. (Note: The three pitch types include fastballs, “hard” cutters, and sinkers, which are all meant to be thrown as hard as possible, as opposed to, for instance, a change up or a curve ball). Interactions between pitch type and pitch count, pitch type and baserunners, and elevation and temperature were also included in this model. Finally, a random in intercept was included for all pitchers to establish a baseline velocity for each. This model can be expressed as:

$$v_{tj} = \delta_0 + \delta X_{tj} + d_j + \epsilon_{tj}$$

$$\epsilon_{tj} \sim N(0, \sigma_{epsilon})$$

$$d_j \sim N(0, \sigma_d)$$

where δ_0 is the fixed intercept of the model, δ is a vector of fixed effects,

d_j is the random effect for the j-th pitcher, v_{tj}, X_{tj}, ϵ_{tj} are the velocity, a vector of covariates, and the residual, respectively, of the t-th pitch for the j-th pitcher and $j = 1 \cdots J$ and $t = 1 \cdots T_j$.

ϵ_{tj} can be interpreted as the velocity above or below expectation for pitcher j (i.e. when $\epsilon_{tj} = 0$ this indicates that the velocity of pitch t for pitcher j is exactly average given that pitcher and set of external factors). This string of residuals will be searched to look for stretches that are more positive that usual (i.e. a "hot" streak) or more negative than usual (i.e a "cold" streak). This search will be accomplished by viewing these residuals ($epsilon_{tj}$) as the emissions from a hidden Markov model.

3.2 MHMM

This project is interested in looking for "hot" and 'cold" states of pitchers, therefore, a two-state hidden Markov model is chosen to look for evidence of these two proposed states. Now, a separate two state hidden Markov model could be fit for each pitcher j yielding J different models, one for each pitcher. Rather, one model will be fit using all the data at once with a random effect to account for differences between pitchers.

This type of model, termed a mixed hidden Markov model (MHMM), is described in Altman (2007).

$$\epsilon_{tj} \sim Normal(\mu_{tj}, \sigma^2)$$

where $\tau^2 = \frac{1}{\sigma^2}$

$$\mu_{tj} = (\beta_1 + b_{j1})\mathcal{I}(S_{tj} = 1) + (\beta_2 + b_{j2})\mathcal{I}(S_{tj} = 2)$$

$$S_{tj} \sim Bern(p_{tj}) + 1$$

$$logit(p_{tj}) = (\gamma_1 + g_{j1})\mathcal{I}(S_{(t-1)j} = 1) + (\gamma_2 + g_{j2})\mathcal{I}(S_{(t-1)j} = 2)$$

$$\beta_1 \sim Uniform(-20, 20)$$

$$\beta_2 \sim Uniform(\beta_1, 20)$$

$$b_{j1} \sim Uniform(-20, 20)$$

$$b_{j2} \sim Uniform(b_{j1} + \beta_1 - \beta_2, 20)$$

$$g_{j1} \sim Uniform(-7, 7); g_{j2} \sim Uniform(-7, 7)$$

$$\gamma_1 \sim Uniform(-7, 7); \gamma_2 \sim Uniform(-7, 7)$$

$$\tau^2 \sim \Gamma(0.001, 0.001);$$

S_{0j} is the initial state for the j-th player and S_{tj} is the state at time t for player j. β_1 and β_2 are the means of the emission distributions for an average pitcher in state 1 and 2, respectively, and β_2 is forced to be larger than β_1. b_{j1} and b_{j2} are the player specific effects for the means of the emissions in states 1 and 2, respectively. Additionally, the model requires that $\beta_2 + b_{j2} \geq \beta_1 + b_{j1}$.

The transition matrix is defined by the γ's and the g's. The transition probabilities are:

$$\nu_{1,2,j} = P(S_{tj} = 2|S_{(t-1)j} = 1) = expit(\gamma_1 + g_{j1})$$

$$\nu_{1,1,j} = P(S_{tj} = 1|S_{(t-1)j} = 1) = 1 - P(S_{tj} = 2|S_{(t-1)j} = 1)$$

$$\nu_{2,2,j} = P(S_{tj} = 2|S_{(t-1)j} = 2) = expit(\gamma_2 + g_{j2})$$

$$\nu_{2,1,j} = P(S_{tj} = 1|S_{(t-1)j} = 2) = 1 - P(S_{tj} = 2|S_{(t-1)j} = 2)s$$

The transition matrix for the j-th individual is

$$\Gamma_j = \begin{pmatrix} \nu_{1,1,j} & \nu_{1,2,j} \\ \nu_{2,1,j} & \nu_{2,2,j} \end{pmatrix}$$

4 Results

10 chains of 10000 draws each were run for all of the parameters. The first 9000 draws were then thrown out of each chain.

In 2015, 6 of the chains needs to be thrown out because they failed to converge for beta.

4.1 Model Diagnostics

Posterior predictive checks and Gelman convergence plots and trace plots.

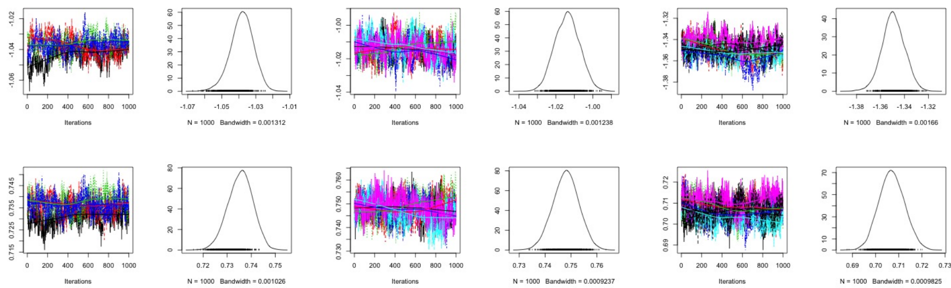


Figure 4: Trace plots of the posterior distribution of β

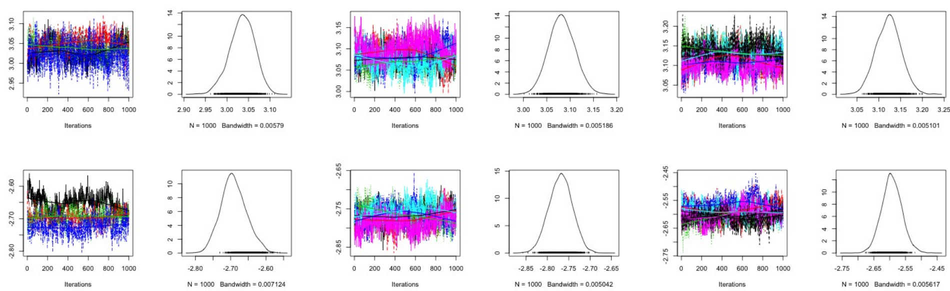


Figure 5: Trace plots of the posterior distribution of γ

In all three years, first 5000 draws we run as a burn-in period and then another 10000 draws were run. Based on inspection of the trace plots 5000 is not enough to reach convergence and and additional 9000 of the 10000 draws were removed. Finally, each chain was checked for convergence visually and chains that still did not appear to converge were removed.

4.1.1 2015

First 5000 draws we run as a burn-in period and then another 10000 draws were run. Based on inspection of the trace plots 5000 is not enough to reach convergence and and additional 9000 of the 10000 draws were removed. Finally, each chain was checked for convergence visually and chains that still did not appear to converge were removed. In 2015 four chains were left. The gelman plot for 2015 can be seen and the Gelman diagnostic for the last 1000 draws of 4 chains is beta1 = 1.36 (upper CI 1.86), beta2 = 1.33 (upper CI 1.79). Gamma1 = 1.22 (1.58) and gamma2 = 1.86 (2.88). Tau = 1.03 (1.1).

4.1.2 2016

In 2016 four chains were removed. The gelman plot for 2016 can be seen and the Gelman diagnostic for the last 1000 draws of 6 chains is beta1 = 1.07 (upper CI 1.16), beta2 = 1.13 (upper CI 1.29). Gamma1 = 1.19 (1.42) and gamma2 = 1.14 (1.32). Tau = 1.04 (1.09).

4.1.3 2017

In 2017 three chains were removed. The gelman plot for 2017 can be seen and the Gelman diagnostic for the last 1000 draws of 7 chains is beta1 = 1.22 (1.48), beta2 = 1.17 (1.36). Gamma1 = 1.21 (1.44) and gamma2 = 1.25 (1.51). Tau = 1.03 (1.06).

This paper suggests that the Gelman diagnostic should be below 1.2 for all parameters to demonstrate convrgence. We have some problems with that in 2015. Brooks, S. P., and A. Gelman. 1997. General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics 7: 434455.

Gelman, A., and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science 7: 457511.

	To Hot	To Cold
From Hot	0.95413544	0.04586456
From Cold	0.06349286	0.93650714

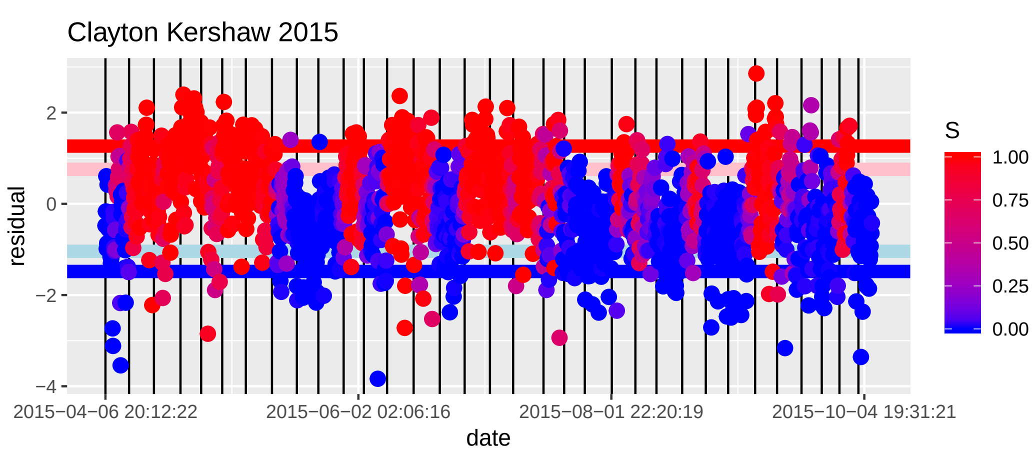
Figure 6: 2015 Transition Matrix

	To Hot	To Cold
From Hot	0.95615461	0.04384539
From Cold	0.05902627	0.94097373

Figure 7: 2016 Transition Matrix

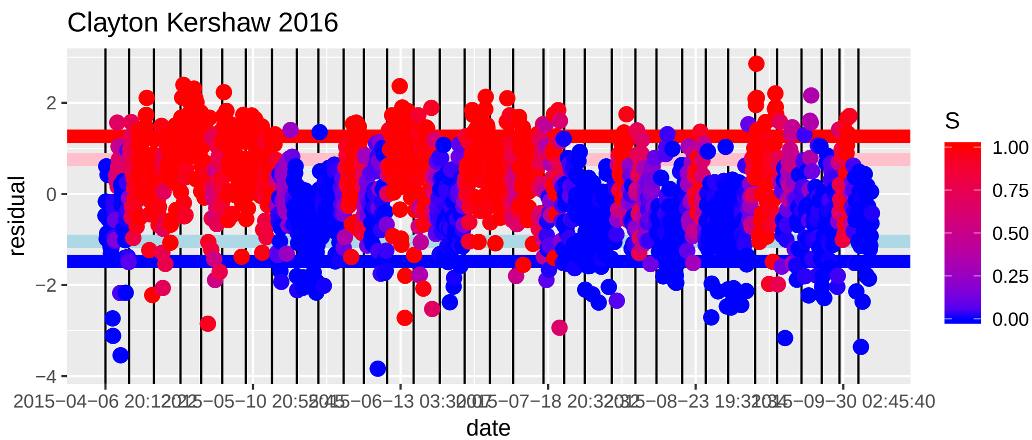
4.2 Parameter Estimates

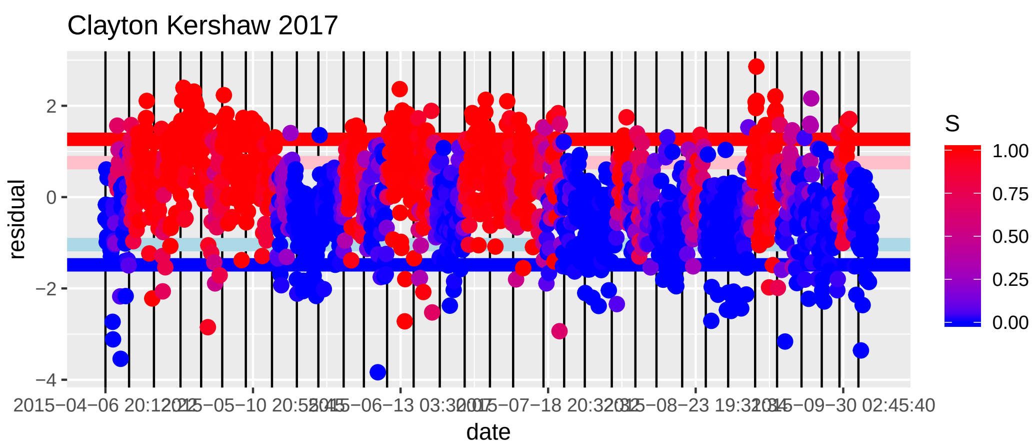
Year	β_1	β_2	γ_1	γ_2	τ
2015	-1.038 (-1.052, -1.026)	0.736 (0.725, 0.745)	3.035 (2.975, 3.088)	-2.691 (-2.758, -2.616)	0.947 (0.940, 0.954)
2016	-1.014 (-1.0266, -1.0009)	0.748 (0.7381, 0.7575)	3.082 (3.029, 3.138)	-2.769 (-2.822, -2.717)	0.9109 (0.9039, 0.9179)
2017	-1.349 (-1.3681, -1.3307)	0.707 (0.6967, 0.7178)	3.122 (3.067, 3.178)	-2.593 (-2.655, -2.529) (	0.8938 (0.8869, 0.9010)



	To Hot	To Cold
From Hot	0.95780989	0.04219011
From Cold	0.06956167	0.93043833

Figure 8: 2017 Transition Matrix





4.3 Prediction

Once the model has been fit, posterior means of each of the parameters can be used as a point estimate. The betas, gammas, and tau are parameters that apply to all players and b and g are parameters that are player specific. Using all of these point estimates, a Hidden Markov model can be constructed based on the posterior estimates.

Prediction with Viterbi2.

5 Conclusions and Future Work

The lag doesn't seem long enough to really fit the data well. A more complex model is probably needed. We view this paper as a starting point. Buchholz is an interesting example in 2016.

Appendix

2015 Model Fixed Effects

	Estimate	Std. Error	t value
(Intercept)	92.65	0.2018	459
pi_pitch_typeFC	-2.429	0.03261	-74.47
pi_pitch_typeSI	-0.3927	0.01311	-29.95
pitch_count	-0.002428	0.0001333	-18.22
baserunnersTRUE	0.2647	0.00817	32.39
elevation	-7.142e-05	2.121e-05	-3.368
gtemp	0.002234	0.0003488	6.404
pi_pitch_typeFC:pitch_count	-0.003438	0.0004977	-6.907
pi_pitch_typeSI:pitch_count	-0.002411	0.0002105	-11.45
pi_pitch_typeFC:baserunnersTRUE	-0.1494	0.0286	-5.223
pi_pitch_typeSI:baserunnersTRUE	-0.002739	0.01264	-0.2167
elevation:gtemp	8.037e-07	2.885e-07	2.786

2016 Model Fixed Effects

2017 Model Fixed Effects

	Estimate	Std. Error	t value
(Intercept)	92.43	0.1932	478.4
pi_pitch_typeFC	-2.935	0.04205	-69.8
pi_pitch_typeSI	-0.3944	0.01367	-28.86
pitch_count	-0.001894	0.000136	-13.93
baserunnersTRUE	0.2234	0.008291	26.95
elevation	1.33e-05	2.277e-05	0.5843
gtemp	0.006481	0.0003472	18.67
pi_pitch_typeFC:pitch_count	-0.00587	0.0006511	-9.016
pi_pitch_typeSI:pitch_count	-0.002994	0.0002215	-13.52
pi_pitch_typeFC:baserunnersTRUE	0.01535	0.03731	0.4114
pi_pitch_typeSI:baserunnersTRUE	0.0503	0.0132	3.81
elevation:gtemp	-2.404e-07	2.981e-07	-0.8064

	Estimate	Std. Error	t value
(Intercept)	93.04	0.1779	523.1
pi_pitch_typeFC	-2.96	0.04273	-69.26
pi_pitch_typeSI	-0.6632	0.01424	-46.56
pitch_count	-0.00335	0.0001399	-23.94
baserunnersTRUE	0.1913	0.008494	22.52
elevation	4.565e-05	2.008e-05	2.273
gtemp	0.001481	0.0003866	3.831
pi_pitch_typeFC:pitch_count	-0.004052	0.0006807	-5.953
pi_pitch_typeSI:pitch_count	-0.001141	0.0002302	-4.959
pi_pitch_typeFC:baserunnersTRUE	0.06236	0.039	1.599
pi_pitch_typeSI:baserunnersTRUE	0.1241	0.01366	9.079
elevation:gtemp	-3.201e-07	2.633e-07	-1.216

References

Altman, R. M. (2007): "Mixed hidden markov models: An extension of the hidden markov model to the longitudinal data setting," *Journal of the American Statistical Association*, 102.

Arkes, J. (2010): "Revisiting the hot hand theory with free throw data in a multivariate framework," *Journal of Quantitative Analysis in Sports*, 6.

Arkes, J. (2013): "Misses in "hot hand" research," *Journal of Sports Economics*, 14, 401–410.

Bocskocsky, E. J., A. and S. Carolyn (2014): "The hot hand: A new approach to an "old fallacy"," *Sloan Sports Conference*.

Gilovich, V. R., T. and A. Tversky (1985): "The hot hand in basketball: On the misperception of random sequences," *Cognitive Psychology*, 17, 295–314.

Huizinga, J. and S. Weil (2009): "Hot hand or hot head: The truth about heat checks in the nba," *MIT Sloan Sports Analytics Conference*.

Humphreys, K. (1998): "The latent markov chain with multivariate random effects," *Sociological Methods and Research*, 26, 269–299.

Koehler, J. and C. Conley (2003): "The "hot hand" myth in professional basketball," *Journal of Sport & Exercise Psychology*, 25, 253–259.

Maruotti, A. (2011): "Mixed hidden markov models for longitudinal data: An overview," *International Statistical Review*, 79, 427–454.

Meeden, G. and S. Vardeman (1997): "A simple hidden markov model for bayesian modeling with time dependent data," .

Seltman, H. (2002): "Hidden markov models for the analysis of biological rhythm data," in *Case Studies in Bayesian Statistics*, New York: Springer, chapter 5, 397–405.

Shirley, K., D. Small, K. Lynch, S. Maisto, and D. Oslin (2010): "Hidden markov models for alcoholism treatment trial data," *The Annals of Applied Statistics*, 4, 366–395.

Tversky, A. and T. Gilovich (1989): "The "hot hand": Statistical reality or cognitive illusion?" *Chance*, 2, 31–34.