Hot or Not?
A Nonparametric Formulation of the Hot Hand in Baseball

Amanda Glazer
amandaglazer@berkeley.edu
Joint work with Lisa Goldberg

What is the hot hand?

A player that has experienced recent success is more likely to continue to do so than one that has not.

Robert Hooke (1989) on the hot hand

"In almost every competitive activity in which I've ever engaged (baseball, basketball, golf, tennis, even duplicate bridge), a little success generates in me a feeling of confidence which, as long as it lasts, makes me do better than usual. Even more obviously, a few failures can destroy this confidence, after which for a while I can't do anything right"

LeBron on the "Hot Hand Farce"

"I guarantee the analytics people has never ever been in the zone in their life."

The original hot hand study

Gilovich, Vallone and Tversky (1985)

Do players hit a higher percentage of their shots after just having made the last k shots, than having just missed the last k shots?

$$\hat{P}^i(\text{hit}|\text{k hits}) - \hat{P}^i(\text{hit}|\text{k misses})$$

Found no evidence of the hot hand

Correct result? Endogeneity? Small sample bias (Miller and Sanjurjo, 2018)?

Small Sample Bias

Three-flip sequence	Proportion of Hs on recorded flips
TTT	–
TTH	–
TH<u>T</u>	0
H<u>T</u>T	0
TH<u>H</u>	1
H<u>T</u>H	0
H<u>H</u>T	$\frac{1}{2}$
H<u>HH</u>	1
Expectation:	$\frac{5}{12}$

Do the Golden State Warriors have hot hands?

Daks, Desai and Goldberg (2018)

Permutation tests with the Gilovich, Vallone and Tversky test statistic

No evidence of a hot hand for Steph Curry, Klay Thompson and Kevin Durant

Previous Approaches in Baseball

Most approaches have not found evidence of a hot hand in baseball (Bar-Eli et al. 2006)

Approaches that have found evidence argue that previous research has had low power and players should be grouped to increase power (Stern 1995, Green and Zwiebel 2016)

ILLUSTRATION BY KELSEY DAKE

1.00–OBJECTIVES OF THE GAME

1.01 Baseball is a game between two teams of nine players each, under direction of a manager, played on an enclosed field in accordance with these rules, under jurisdiction of one or more umpires.

1.02 The offensive team's objective is to have its batter become a runner, and its runners advance.

Key Terms

Plate appearance (PA) = a batter's turn at the plate

On-base percentage (OBP) = how frequently a batter reaches base (hits, walks, and times hit by pitch) per plate appearance

Defining the batter hot hand

1. Does a batter perform better if they have performed well in their last L plate appearances, outside of the effects of all other factors?

2. **Does fan's *perception* of the hot hand, batters that have performed well recently will continue to do so, exist?**

Green and Zwiebel 2016

State: Average of outcome, Y, for last L PAs

Ability: Average of outcome, Y, for all PAs except the 50 before and 50 after

Model:

$$Y_{it} = \alpha + \underbrace{\gamma \cdot \text{State}_{it}}_{\text{Hot-Hand Effect}} + \delta \cdot \text{Ability}_{it} + \beta \cdot X_{it} + \epsilon_{it}$$

Consider five outcomes: hit, homerun, strikeout, on base, walk

Green and Zwiebel 2016

Results:

On Base Batter

	(1) OLS	(2) Logit	(3) OLS_dist	(4) Prop5	(5) Add5	(6) Dist5
main						
state	0.0658*** (14.61)	0.292*** (14.76)	0.0114*** (8.63)			
hot				0.0164*** (7.95)	0.0169*** (8.61)	0.00762*** (4.45)
cold				-0.0133*** (-7.35)	-0.0134*** (-7.43)	-0.00958*** (-4.81)
batter_ability	0.551*** (20.92)	2.440*** (22.36)	0.615*** (20.37)	0.602*** (21.30)	0.602*** (21.31)	0.603*** (20.01)
pitcher_ability	0.578*** (44.25)	2.597*** (44.18)	0.576*** (42.86)	0.580*** (44.34)	0.580*** (44.31)	0.577*** (43.12)
samehand	-0.0245*** (-19.49)	-0.110*** (-19.60)	-0.0247*** (-18.94)	-0.0246*** (-19.32)	-0.0246*** (-19.33)	-0.0249*** (-18.94)
batter_home	0.0127*** (16.08)	0.0569*** (16.13)	0.0129*** (15.82)	0.0128*** (16.12)	0.0128*** (16.12)	0.0129*** (15.86)
Observations	1489346	1489346	1382576	1489346	1489346	1399046

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Our Data

Major League Baseball (MLB) data from retrosheet.org

All teams (30) from the 2018 season

All players with more than 100 plate appearances (PAs)

Choice of Test Statistic

Correlation between lag L OBP and whether the current PA results in the player making it on base, for L = 5, 10, 25

Autocorrelation

Regression coefficient

Gilovich, Vallone and Tversky test statistic

Permutation Tests

0 0 0 0 0 0 1 0 0 0 1 1 1 0 1 0 1 1 0 0 1 1 0 0 0 1 1 1 0 1 1 0 1 0 1 0 1 0 1 1 0 0 1 1 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 1 0 1 1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 1 0 0 0 1 1 0 0 1 1 0 1 0 0 0 0 0 0 1

Assumption: If the hot hand exists, our original test statistic should be extreme compared to random shufflings of the data

Our Methodology: Permutation Tests

For each player:

1. Calculate the test statistic (e.g., correlation between state and next PA outcome) for the sequence of PAs
2. Shuffle the PAs 10000 times
3. For each shuffling, calculate the test statistic
4. P-value = proportion of shufflings that result in a test statistic greater than or equal to our original test statistic

Permutation Tests Pros and Cons

Pros:

Minimal assumptions

Conceptually clear

Cons:

Conservative (can have low power)

Choice of Lag

Consider 24 1s followed by 24 0s for the whole season:

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1.....

What is the correlation between lag 25 OBP and whether you make it OB the next PA?

Choice of Lag

Consider 24 1s followed by 24 0s for the whole season:

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1.....

What is the correlation between lag 25 OBP and whether you make it OB the next PA? **-0.136**

In general: lag longer than streak → negative correlation

Pooling data

What happens when we feed the model random data?

1. Create 400 players with OBP ranging from .25 to .45
2. Generate their PAs from Binom(500, OBP)

What percent of the time will the Green and Zwiebel model yield significance of the state variable (at 0.05 level)?

Pooling data

What happens when we feed the model random data?

1. Create 400 players with OBP ranging from .25 to .45
2. Generate their PAs from Binom(500, OBP)

What percent of the time will the Green and Zwiebel model yield significance of the state variable (at 0.05 level)? **99% of the time**
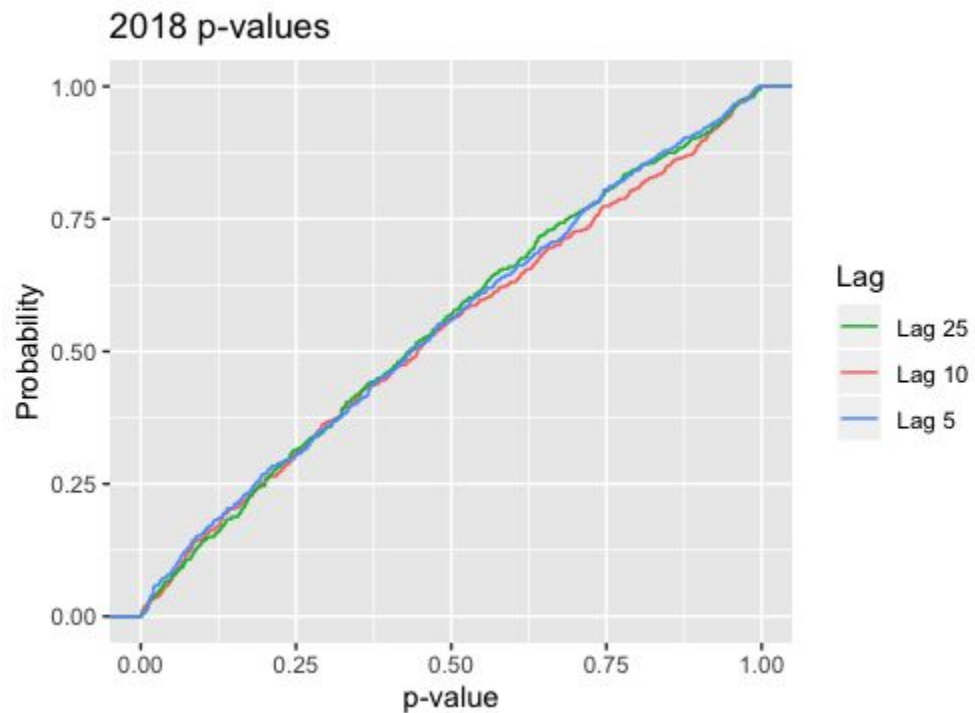
Power

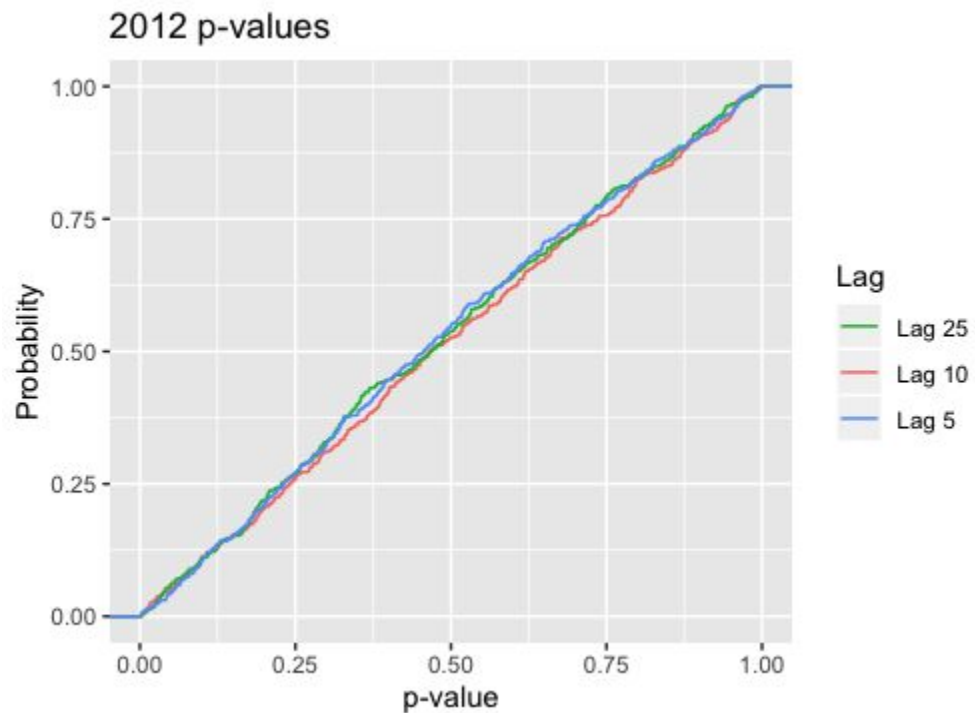Two-state markov chain with transition probability 0.05

Hot/Cold state OBP

	0.6/0.2	0.55/0.25	0.5/0.3
Lag 5 Correlation	1	0.88	0.32
Lag 10 Correlation	0.99	0.85	0.34
Autocorrelation	0.9	0.51	0.18
Tversky Statistic	0.97	0.65	0.23

Our Results

2018 p-values

Our Results

2012 p-values

Other nonparametric formulations

Mike Trout 2019: Hot or Not?
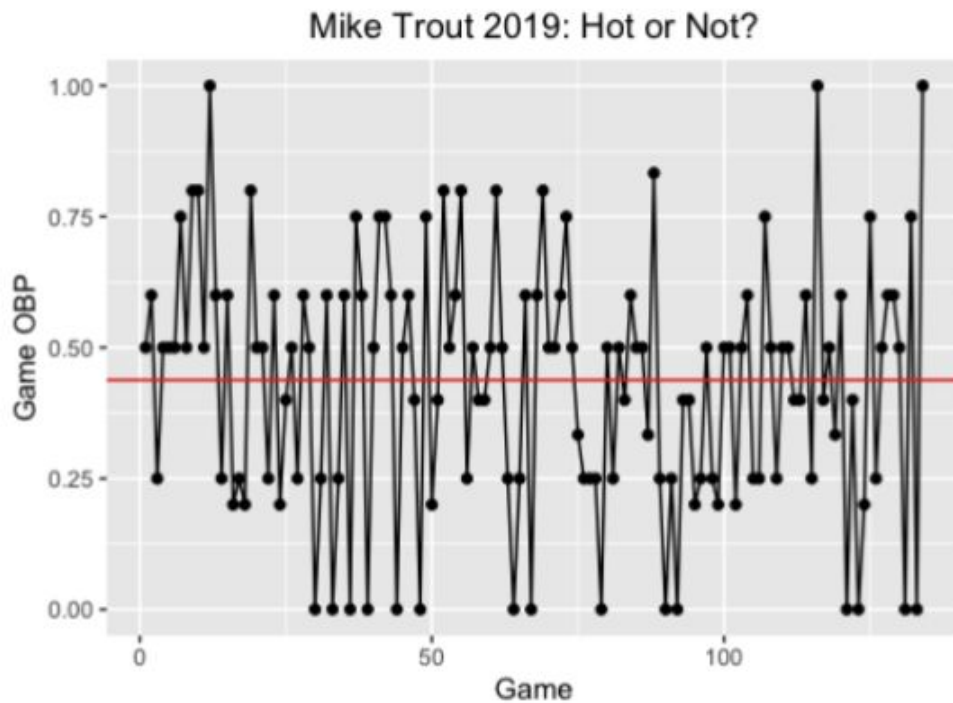
Summary

Our nonparametric tests show no evidence of the batter hot hand in baseball

Nonparametric tests are conceptually clear with minimal assumptions

Pooling data can yield high type I error rates

    Trade-off between type I error rate and power