

The Hot Hand Fallacy: Cognitive Mistake or Equilibrium Adjustments? Evidence from Major League Baseball

Brett Green – UC Berkeley
Jeffrey Zwiebel – Stanford University

August 2015

- The “hot hand fallacy” is frequently invoked as a motivating example for behavioral finance and economics.
- Looking carefully at the literature, we were not convinced there is a fallacy.
- Led us to some more general thoughts about streakiness, equilibrium reactions, and when we should expect to find a hot hand (if it exists).
 - Also relevant for whether we should expect to find evidence of “skill”

Outline for Talk

- 1) Background on the hot hand
- 2) Our hypothesis - Endogeneity and baseball vs. basketball
- 3) Our findings
 - Evidence of a hot hand in baseball.
 - Simulations to give sense of an underlying process matching our results.
 - Do teams correctly estimate/respond to the hot hand?

Background

Starting Point: Gilovich, Vallone and Tversky (GVT) (1985)

- 91% of fans believe a [basketball] player has “a better chance” of making a shot after having just made his last two or three shots than he does after having just missed his last two or three shots.
- Belief is pervasive across players, coaches, announcers and analysts.
- GVT argue that the data does not support these beliefs.

TABLE 1

Probability of Making a Shot Conditioned on the Outcome of Previous Shots for Nine Members of the Philadelphia 76ers

Player	<i>P</i> (hit/3 misses)	<i>P</i> (hit/2 misses)	<i>P</i> (hit/1 miss)	<i>P</i> (hit)	<i>P</i> (hit/1 hit)	<i>P</i> (hit/2 hits)	<i>P</i> (hit/3 hits)	Serial correlation <i>r</i>
Clint Richardson	.50 (12)	.47 (32)	.56 (101)	.50 (248)	.49 (105)	.50 (46)	.48 (21)	-.020
Julius Erving	.52 (90)	.51 (191)	.51 (408)	.52 (884)	.53 (428)	.52 (211)	.48 (97)	.016
Lionel Hollins	.50 (40)	.49 (92)	.46 (200)	.46 (419)	.46 (171)	.46 (65)	.32 (25)	-.004
Maurice Cheeks	.77 (13)	.60 (38)	.60 (126)	.56 (339)	.55 (166)	.54 (76)	.59 (32)	-.038
Caldwell Jones	.50 (20)	.48 (48)	.47 (117)	.47 (272)	.45 (108)	.43 (37)	.27 (11)	-.016
Andrew Toney	.52 (33)	.53 (90)	.51 (216)	.46 (451)	.43 (190)	.40 (77)	.34 (29)	-.083
Bobby Jones	.61 (23)	.58 (66)	.58 (179)	.54 (433)	.53 (207)	.47 (96)	.53 (36)	-.049
Steve Mix	.70 (20)	.56 (54)	.52 (147)	.52 (351)	.51 (163)	.48 (77)	.36 (33)	-.015
Daryl Dawkins	.88 (8)	.73 (33)	.71 (136)	.62 (403)	.57 (222)	.58 (111)	.51 (55)	-.142**
Weighted means	.56	.53	.54	.52	.51	.50	.46	-.039

Background

A large literature following GVT, focused mainly on basketball, but including studies of other sports as well, mostly agrees with GVT.

- Finds little evidence of a hot hand in sports
- Instead, beliefs in the hot hand are interpreted as a manifestation of a fundamental cognitive mistake
 - People infer patterns from random data.
- Furthermore, this literature argues that this misperception leads to sub-optimal decision making

This “hot hand fallacy” is frequently invoked as motivation for behavioral economics and behavioral finance

Background

A large literature following GVT, focused mainly on basketball, but including studies of other sports as well, mostly agrees with GVT.

- Finds little evidence of a hot hand in sports
- Instead, beliefs in the hot hand are interpreted as a manifestation of a fundamental cognitive mistake
 - People infer patterns from random data.
- Furthermore, this literature argues that this misperception leads to sub-optimal decision making

This “hot hand fallacy” is frequently invoked as motivation for behavioral economics and behavioral finance

Drawbacks of previous research

- 1) Equilibrium effects often ignored
 - Defensive resources are fungible in basketball
 - Offense optimally responds to defense
 - Equilibrium: resources allocated across agents to equate margins
 - * If adjustments are frictionless, we should not expect to find streakiness in the data even if a hot hand exists
- 2) Most tests lack power to detect “reasonable” models of streaky behavior
- 3) Biased estimates of player ability
 - e.g., “The Book” (Tango et al, 2007)

Drawbacks of previous research

- 1) Equilibrium effects often ignored
 - Defensive resources are fungible in basketball
 - Offense optimally responds to defense
 - Equilibrium: resources allocated across agents to equate margins
 - * If adjustments are frictionless, we should not expect to find streakiness in the data even if a hot hand exists
- 2) Most tests lack power to detect “reasonable” models of streaky behavior
- 3) Biased estimates of player ability
 - e.g., “The Book” (Tango et al, 2007)

GVT were aware of defensive responses

TABLE 3
Probability of Making a Second Free Throw Conditioned on the Outcome of the First Free Throw for Nine Members of the Boston Celtics during the 1980–1981 and 1981–1982 Seasons

Player	$P(H_2/M_1)$	$P(H_2/H_1)$	Serial correlation r
Larry Bird	.91 (53)	.88 (285)	-.032
Cedric Maxwell	.76 (128)	.81 (302)	.061
Robert Parish	.72 (105)	.77 (213)	.056
Nate Archibald	.82 (76)	.83 (245)	.014
Chris Ford	.77 (22)	.71 (51)	-.069
Kevin McHale	.59 (49)	.73 (128)	.130
M. L. Carr	.81 (26)	.68 (57)	-.128
Rick Robey	.61 (80)	.59 (91)	-.019
Gerald Henderson	.78 (37)	.76 (101)	-.022

- However, this test (and others like it) are severely underpowered.
 - Need roughly two orders of magnitude more data.

TABLE 3
Probability of Making a Second Free Throw Conditioned on the Outcome of the First Free Throw for Nine Members of the Boston Celtics during the 1980–1981 and 1981–1982 Seasons

Player	$P(H_2/M_1)$	$P(H_2/H_1)$	Serial correlation r
Larry Bird	.91 (53)	.88 (285)	-.032
Cedric Maxwell	.76 (128)	.81 (302)	.061
Robert Parish	.72 (105)	.77 (213)	.056
Nate Archibald	.82 (76)	.83 (245)	.014
Chris Ford	.77 (22)	.71 (51)	-.069
Kevin McHale	.59 (49)	.73 (128)	.130
M. L. Carr	.81 (26)	.68 (57)	-.128
Rick Robey	.61 (80)	.59 (91)	-.019
Gerald Henderson	.78 (37)	.76 (101)	-.022

- However, this test (and others like it) are severely underpowered.
 - Need roughly two orders of magnitude more data.

Illustration of the power issue

Model	(1)	(2)	(3)	(4)	(5)
Pr(H)	70%	70%	70%	70%	70%
P(H Hot)	75%	80%	90%	80%	80%
Pr(H Cold)	65%	60%	50%	60%	70%
Half Life (# of Shots)	5	5	5	5	10
Percent of time hot	10%	10%	10%	20%	10%
Percent of time cold	10%	10%	10%	20%	10%

Results

Pr(H ₂ H ₁)	69.98%	70.19%	70.89%	70.43%	70.19%
Pr(H ₂ M ₁)	70.01%	69.52%	67.89%	69.02%	69.62%
Percent Rejected (5%)	8.26%	10.06%	18.35%	12.66%	11.07%
Percent Rejected (1%)	2.35%	3.29%	7.61%	4.25%	3.90%

Obs per simulation 430

Number of simulations 10000

This project

We argue that endogenous adjustments should equate margins across players in basketball; not so in baseball.

- Hitters are faced sequentially.
- Little scope for defensive adjustments that transfer resources from one hitter to another.
- Equating margins across hitters would not necessarily be optimal even if it was possible.

Hence, we revisit the issue using panel data from major league baseball.

- Little concern about endogeneity and virtually unlimited data.
- Is there evidence of a hot hand in baseball?

We can also measure defensive responses to recent performance.

- Do teams respond to streaky behavior?
- Is their response consistent with drawing correct inference about the magnitude of the hot-hand effect?

This project

We argue that endogenous adjustments should equate margins across players in basketball; not so in baseball.

- Hitters are faced sequentially.
- Little scope for defensive adjustments that transfer resources from one hitter to another.
- Equating margins across hitters would not necessarily be optimal even if it was possible.

Hence, we revisit the issue using panel data from major league baseball.

- Little concern about endogeneity and virtually unlimited data.
- Is there evidence of a hot hand in baseball?

We can also measure defensive responses to recent performance.

- Do teams respond to streaky behavior?
- Is their response consistent with drawing correct inference about the magnitude of the hot-hand effect?

Our Findings

- 1) Is there evidence of a hot hand in major league baseball?
 - We find strong evidence of a hot-hand effect across all categories that we analyze.
 - The magnitudes are significant – both statistically and strategically
 - * Difference between “cold” and “hot” batter corresponds to roughly a one quartile increase (e.g., from 50th percentile to 75th percentile).
 - * Simulations suggest the underlying data-generating process exhibits significantly larger magnitudes.
- 2) Do defenses respond to streaky performance?
 - Opposing teams clearly respond to “streaky” batters by “pitching around” them resulting in more walks.
- 3) Is their response consistent with correctly inferring the magnitude of the hot-hand effect?
 - They seem to draw correct inference about the magnitude of the hot-hand effect based on performance in the prior 25 outcomes.
 - But overestimate its magnitude based on very recent performance.

Our Findings

- 1) Is there evidence of a hot hand in major league baseball?
 - We find strong evidence of a hot-hand effect across all categories that we analyze.
 - The magnitudes are significant – both statistically and strategically
 - * Difference between “cold” and “hot” batter corresponds to roughly a one quartile increase (e.g., from 50th percentile to 75th percentile).
 - * Simulations suggest the underlying data-generating process exhibits significantly larger magnitudes.
- 2) Do defenses respond to streaky performance?
 - Opposing teams clearly respond to “streaky” batters by “pitching around” them resulting in more walks.
- 3) Is their response consistent with correctly inferring the magnitude of the hot-hand effect?
 - They seem to draw correct inference about the magnitude of the hot-hand effect based on performance in the prior 25 outcomes.
 - But overestimate its magnitude based on very recent performance.

Our Findings

- 1) Is there evidence of a hot hand in major league baseball?
 - We find strong evidence of a hot-hand effect across all categories that we analyze.
 - The magnitudes are significant – both statistically and strategically
 - * Difference between “cold” and “hot” batter corresponds to roughly a one quartile increase (e.g., from 50th percentile to 75th percentile).
 - * Simulations suggest the underlying data-generating process exhibits significantly larger magnitudes.
- 2) Do defenses respond to streaky performance?
 - Opposing teams clearly respond to “streaky” batters by “pitching around” them resulting in more walks.
- 3) Is their response consistent with correctly inferring the magnitude of the hot-hand effect?
 - They seem to draw correct inference about the magnitude of the hot-hand effect based on performance in the prior 25 outcomes.
 - But overestimate its magnitude based on very recent performance.

Our Findings

- 1) Is there evidence of a hot hand in major league baseball?
 - We find strong evidence of a hot-hand effect across all categories that we analyze.
 - The magnitudes are significant – both statistically and strategically
 - * Difference between “cold” and “hot” batter corresponds to roughly a one quartile increase (e.g., from 50th percentile to 75th percentile).
 - * Simulations suggest the underlying data-generating process exhibits significantly larger magnitudes.
- 2) Do defenses respond to streaky performance?
 - Opposing teams clearly respond to “streaky” batters by “pitching around” them resulting in more walks.
- 3) Is their response consistent with correctly inferring the magnitude of the hot-hand effect?
 - They seem to draw correct inference about the magnitude of the hot-hand effect based on performance in the prior 25 outcomes.
 - But overestimate its magnitude based on very recent performance.

Endogenous responses: Basketball vs Baseball

Basketball: Offensive players simultaneously guarded by defense.

- Defenses are fluid
 - Double teams, man vs zone, rotations
 - Continuously adjusted
- (Roughly) optimal offensive strategy: Player A shoots iff

$$\mathbb{P}(A \text{ makes} | \text{Defense}) > \mathbb{P}(B \text{ makes} | \text{Defense}).$$

- Consequently, optimal defensive strategy: Shift defensive resources to MinMax probability over offensive players - i.e., equate margins.

In equilibrium

$$\mathbb{P}(A \text{ makes} | \text{Defense}) = \mathbb{P}(B \text{ makes} | \text{Defense}).$$

Endogenous responses: Basketball vs Baseball

Basketball: Offensive players simultaneously guarded by defense.

- Defenses are fluid
 - Double teams, man vs zone, rotations
 - Continuously adjusted
- (Roughly) optimal offensive strategy: Player A shoots iff

$$\mathbb{P}(A \text{ makes} | \text{Defense}) > \mathbb{P}(B \text{ makes} | \text{Defense}).$$

- Consequently, optimal defensive strategy: Shift defensive resources to MinMax probability over offensive players - i.e., equate margins.

In equilibrium

$$\mathbb{P}(A \text{ makes} | \text{Defense}) = \mathbb{P}(B \text{ makes} | \text{Defense}).$$

Endogenous responses: Basketball vs Baseball

Baseball: The defense faces each batter sequentially as opposed to simultaneously

- Defense should optimize batter by batter
- Little scope for transferring resources across batters
- Also, no reason such transfers would be optimal anyway
 - Batter order is fixed: no analogous offensive adjustment in baseball.
- One significant (unobservable) adjustment: “pitch around”
 - Doing so is costly: will \uparrow likelihood of a walk.
 - We will use this to test for endogenous response.
 - Walks do not negatively influence our statistics.

Evidence: Basketball shooting percentage relatively uninformative about player ability compared to batting average

Endogenous responses: Basketball vs Baseball

Baseball: The defense faces each batter sequentially as opposed to simultaneously

- Defense should optimize batter by batter
- Little scope for transferring resources across batters
- Also, no reason such transfers would be optimal anyway
 - Batter order is fixed: no analogous offensive adjustment in baseball.
- One significant (unobservable) adjustment: “pitch around”
 - Doing so is costly: will \uparrow likelihood of a walk.
 - We will use this to test for endogenous response.
 - Walks do not negatively influence our statistics.

Evidence: Basketball shooting percentage relatively uninformative about player ability compared to batting average

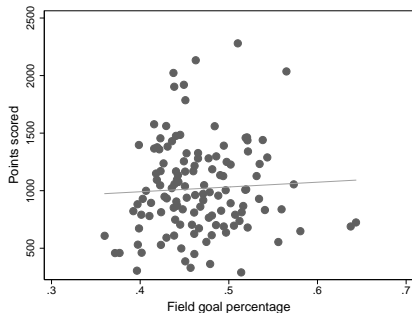
Endogenous responses: Basketball vs Baseball

Baseball: The defense faces each batter sequentially as opposed to simultaneously

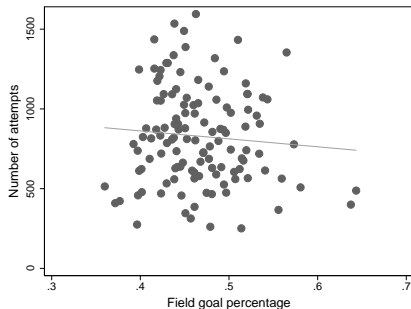
- Defense should optimize batter by batter
- Little scope for transferring resources across batters
- Also, no reason such transfers would be optimal anyway
 - Batter order is fixed: no analogous offensive adjustment in baseball.
- One significant (unobservable) adjustment: “pitch around”
 - Doing so is costly: will \uparrow likelihood of a walk.
 - We will use this to test for endogenous response.
 - Walks do not negatively influence our statistics.

Evidence: Basketball shooting percentage relatively uninformative about player ability compared to batting average

Basketball



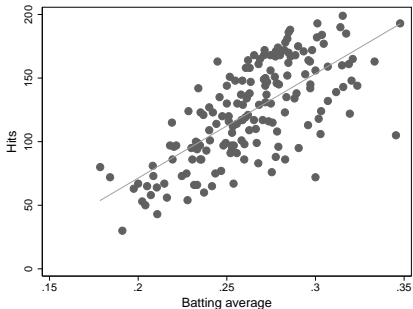
(a) Successes



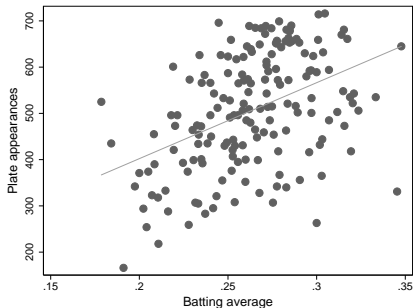
(b) Attempts

- Measures of “ability” and shooting percentage are not positively correlated.
 - Shooting percentage of top scorers, all-stars, etc. is statistically indistinguishable from average starter.

Baseball



(c) Successes



(d) Attempts

- Measures of “ability” and batting average are strongly correlated.

Our General Hypothesis

- Skilled activity will generally demonstrate both persistent long-run variation in ability and transitory short-run variation in ability – i.e., streakiness.
- In some activities (e.g., basketball), endogenous responses will equate margins and thereby confound evidence of a hot hand.
 - Outcomes may not exhibit serial correlation even if the underlying ability process does.
- One simple indication that such an effect exists: it should equate margins for both long-run and short-run variation in ability
 - If and only if there is persistent variation in long-run performance, should we also expect to find evidence for a hot hand.

Our General Hypothesis

- Skilled activity will generally demonstrate both persistent long-run variation in ability and transitory short-run variation in ability – i.e., streakiness.
- In some activities (e.g., basketball), endogenous responses will equate margins and thereby confound evidence of a hot hand.
 - Outcomes may not exhibit serial correlation even if the underlying ability process does.
- One simple indication that such an effect exists: it should equate margins for both long-run and short-run variation in ability
 - If and only if there is persistent variation in long-run performance, should we also expect to find evidence for a hot hand.

Our General Hypothesis

- Skilled activity will generally demonstrate both persistent long-run variation in ability and transitory short-run variation in ability – i.e., streakiness.
- In some activities (e.g., basketball), endogenous responses will equate margins and thereby confound evidence of a hot hand.
 - Outcomes may not exhibit serial correlation even if the underlying ability process does.
- One simple indication that such an effect exists: it should equate margins for both long-run and short-run variation in ability
 - If and only if there is persistent variation in long-run performance, should we also expect to find evidence for a hot hand.

What is a Hot Hand?

- No clear consensus on a precise definition
- Generally adopted notion of a **hot hand**: short-term predictability
 - How much does recent success predict future success after controlling for other predictive factors?
- One obvious factor to control for is the player's ability.
- But, player ability changes over time...is this a hot hand?
 - Frequency with which ability changes are crucial to the definition.

What is a Hot Hand?

- No clear consensus on a precise definition
- Generally adopted notion of a **hot hand**: short-term predictability
 - How much does recent success predict future success after controlling for other predictive factors?
- One obvious factor to control for is the player's ability.
- But, player ability changes over time...is this a hot hand?
 - Frequency with which ability changes are crucial to the definition.

What is a Hot Hand?

This fits most people's intuition of a hot hand:

- In 1997, Joey Cora batted 0.475 during a 24 game span (roughly 4 weeks). He batted 0.262 during the rest of the season.

But not this:

- In 2011, Chris Davis hit 5 home runs (1 per 40 attempts).
In 2012, he hit 33 home runs (1 per 15).
In 2013, he hit a league leading 53 home runs (1 per 11).

We define **streakiness** as relatively high-frequency variation in a player's success probability after controlling for external factors.

- e.g., daily, weekly or monthly shocks to ability.
- lower frequency variation (e.g., annual) not considered streaky.

What is a Hot Hand?

This fits most people's intuition of a hot hand:

- In 1997, Joey Cora batted 0.475 during a 24 game span (roughly 4 weeks). He batted 0.262 during the rest of the season.

But not this:

- In 2011, Chris Davis hit 5 home runs (1 per 40 attempts).
In 2012, he hit 33 home runs (1 per 15).
In 2013, he hit a league leading 53 home runs (1 per 11).

We define **streakiness** as relatively high-frequency variation in a player's success probability after controlling for external factors.

- e.g., daily, weekly or monthly shocks to ability.
- lower frequency variation (e.g., annual) not considered streaky.

What is a Hot Hand?

This fits most people's intuition of a hot hand:

- In 1997, Joey Cora batted 0.475 during a 24 game span (roughly 4 weeks). He batted 0.262 during the rest of the season.

But not this:

- In 2011, Chris Davis hit 5 home runs (1 per 40 attempts).
In 2012, he hit 33 home runs (1 per 15).
In 2013, he hit a league leading 53 home runs (1 per 11).

We define **streakiness** as relatively high-frequency variation in a player's success probability after controlling for external factors.

- e.g., daily, weekly or monthly shocks to ability.
- lower frequency variation (e.g., annual) not considered streaky.

Conceptual Framework

A player's success probability in any given attempt is determined by three components:

- **Long-term player-specific component (ability):** Raw talent, skill, speed, strength, hand-eye coordination
 - Only exhibits low-frequency variation over time.
- **Short-term player-specific component (state):** Confidence, attitude, adjustments in technique, health
 - Captures temporal variation in ability (if it exists).
- **External factors (controls):** Opponent ability, team strategy, game situation, stadium, platoon effect, etc.

Primary Specification

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

- Estimate for five different outcomes: hit, homerun, strikeout, on base, walk, both for batters and pitchers (10 total)
- Controls (X): Ability of opponent, stadium, platoon effect, inning, score, runners on-base, home or away, etc.
- Standard errors clustered at the batter (or pitcher) level
- We consider variety of different measurements of state and ability.

Measuring short-term ability (state)

- 1) Use the player's recent success rate (e.g., batting average) to estimate his current state. Focus on $L = 25$, also consider $L = 10, 50$.

$$\text{State}_{it} = R_{it}^L \equiv \frac{1}{L} \sum_{k=1}^L Y_{i,t-k}.$$

- 2) Use $\text{State}_{it} = [\text{Hot}_{it}, \text{Cold}_{it}]$, where

- Additive Cutoffs:

$$\text{Hot}_{it} = 1 (R_{it}^L \geq \text{Ability}_{it} + A)$$

$$\text{Cold}_{it} = 1 (R_{it}^L \leq \text{Ability}_{it} - B)$$

- Proportional Cutoffs:

$$\text{Hot}_{it} = 1 (R_{it}^L \geq q \cdot \text{Ability}_{it})$$

$$\text{Cold}_{it} = 1 (R_{it}^L \leq r \cdot \text{Ability}_{it})$$

3) Distribution-based state measure:

$$State_{it}^d = \text{freq}\{R_{it}^L > R_{is}^L, \forall s \notin \mathcal{S}(t)\}$$

where $\mathcal{S}(t) \in \{1, 2\}$ denotes season half of attempt t .

- In words, where i 's recent success at t lies in the distribution of success rates over the opposite half season.

4) Distribution based cutoffs:

$$Hot_{it} = 1 \left(State_{it}^d > 1 - p \right)$$

$$Cold_{it} = 1 \left(State_{it}^d < p \right)$$

for $p \in \{.01, .05, .1\}$

Measuring long-term ability

Ability: We also consider a variety of methods to measure long-term ability

- Current season success rate not including “window”

$$\text{Ability}_{it} = \frac{1}{N_i - 2W} \times \sum_{s \notin [t-W, t+W]} Y_{is}$$

- Player fixed effect
 - For many dgp: leads to downward bias but asymptotically consistent
- Previous season success rate + current season (up to window)
- Previous two seasons + current season (up to window)
- Career success rate not including current season

Data

- From Major League Baseball for the 2000-2011 seasons
 \approx 2M observations
- Obtained from Retrosheet.org
- Each observation is an *event*: any occurrence that changes the state of the game. Includes roughly 90 variables (e.g., batter, pitcher, score, outs, inning, base-runners, fielders, stadium, date, etc.)

variable	mean	sd	p25	p50	p75	p99	N
Batting	.268	.0223	.254	.268	.282	.323	789
Homerun	.0304	.0169	.0178	.0289	.0415	.0728	789
Strikeout	.191	.059	.147	.186	.228	.346	789
OnBase	.339	.0299	.32	.338	.356	.416	789
Walk	.0966	.0302	.0758	.0939	.113	.18	789
PlateAppearances	2035	1726	668	1413	3034	7132	789
AtBats	1830	1543	594	1283	2754	6563	789

Data

- From Major League Baseball for the 2000-2011 seasons
 \approx 2M observations
- Obtained from Retrosheet.org
- Each observation is an *event*: any occurrence that changes the state of the game. Includes roughly 90 variables (e.g., batter, pitcher, score, outs, inning, base-runners, fielders, stadium, date, etc.)

variable	mean	sd	p25	p50	p75	p99	N
Batting	.268	.0223	.254	.268	.282	.323	789
Homerun	.0304	.0169	.0178	.0289	.0415	.0728	789
Strikeout	.191	.059	.147	.186	.228	.346	789
OnBase	.339	.0299	.32	.338	.356	.416	789
Walk	.0966	.0302	.0758	.0939	.113	.18	789
PlateAppearances	2035	1726	668	1413	3034	7132	789
AtBats	1830	1543	594	1283	2754	6563	789

Overview of Hot Hand Evidence

- Positive and strongly significant γ for all categories in virtually all of our specifications and tests:
 - A .400 increase in OBP in the last 25 PA \implies 26 extra OBP points
 - A .400 increase in BA in the last 25 AB \implies 13 extra BA points
- Using 5% thresholds: Difference between a hot and cold player (of the same ability) is roughly one quartile of the distribution i.e.,
 \approx the difference between a 75th and 50th percentile player.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
 - Strongest effect for hitters: home runs, OBP, walks
 - Strongest effect for pitchers: strikeouts, OBP

Overview of Hot Hand Evidence

- Positive and strongly significant γ for all categories in virtually all of our specifications and tests:
 - A .400 increase in OBP in the last 25 PA \implies 26 extra OBP points
 - A .400 increase in BA in the last 25 AB \implies 13 extra BA points
- Using 5% thresholds: Difference between a hot and cold player (of the same ability) is roughly one quartile of the distribution i.e.,
 - \approx the difference between a 75th and 50th percentile player.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
 - Strongest effect for hitters: home runs, OBP, walks
 - Strongest effect for pitchers: strikeouts, OBP

Overview of Hot Hand Evidence

- Positive and strongly significant γ for all categories in virtually all of our specifications and tests:
 - A .400 increase in OBP in the last 25 PA \implies 26 extra OBP points
 - A .400 increase in BA in the last 25 AB \implies 13 extra BA points
- Using 5% thresholds: Difference between a hot and cold player (of the same ability) is roughly one quartile of the distribution i.e.,
 - \approx the difference between a 75th and 50th percentile player.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
 - Strongest effect for hitters: home runs, OBP, walks
 - Strongest effect for pitchers: strikeouts, OBP

Results: On Base Percentage – Batters

	(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

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results: Home Runs – Batters

	(1) OLS	(2) Logit	(3) OLS_dist	(4) Prop5	(5) Add5	(6) Dist5
main						
state	0.0749*** (15.86)	1.945*** (15.83)	0.00911*** (10.93)			
hot				0.00420*** (6.03)	0.00786*** (9.18)	0.00190*** (3.79)
batter_ability	0.702*** (51.22)	19.13*** (37.55)	0.814*** (58.33)	0.769*** (55.89)	0.752*** (54.41)	0.765*** (55.11)
pitcher_ability	0.376*** (20.49)	11.32*** (22.72)	0.378*** (19.66)	0.378*** (20.49)	0.377*** (20.50)	0.378*** (20.50)
samehand	-0.00452*** (-8.92)	-0.124*** (-7.65)	-0.00454*** (-8.59)	-0.00442*** (-8.58)	-0.00446*** (-8.69)	-0.00442*** (-8.58)
batter_home	0.00209*** (5.35)	0.0562*** (4.84)	0.00196*** (4.81)	0.00207*** (5.25)	0.00208*** (5.29)	0.00206*** (5.24)
Observations	1192266	1192266	1101873	1192266	1192266	1192266

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results: Hits for Batters

	(1) OLS	(2) Logit	(3) OLS_dist	(4) Prop5	(5) Add5	(6) Dist5
main						
state	0.0308*** (5.98)	0.155*** (6.00)	0.00534*** (3.44)			
hot				0.00787*** (4.04)	0.00644*** (3.66)	0.00427* (2.38)
cold				-0.00728*** (-3.41)	-0.00921*** (-4.22)	-0.00447* (-2.07)
batter_ability	0.366*** (15.94)	1.839*** (16.17)	0.386*** (14.67)	0.388*** (16.22)	0.389*** (16.32)	0.382*** (14.81)
pitcher_ability	0.529*** (33.80)	2.688*** (33.15)	0.528*** (32.64)	0.530*** (33.80)	0.530*** (33.80)	0.530*** (32.82)
samehand	-0.0131*** (-12.09)	-0.0657*** (-12.05)	-0.0130*** (-11.68)	-0.0130*** (-12.03)	-0.0131*** (-12.03)	-0.0131*** (-11.83)
batter_home	0.00911*** (10.84)	0.0458*** (10.87)	0.00899*** (10.32)	0.00913*** (10.85)	0.00913*** (10.84)	0.00897*** (10.34)
Observations	1192266	1192266	1101873	1192266	1192266	1113087

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results: Strikeouts for Pitchers

	(1) OLS	(2) Logit	(3) OLS_dist	(4) Prop5	(5) Add5	(6) Dist5
main state	0.118*** (21.99)	0.766*** (22.91)	0.0288*** (17.53)			
hot				0.0195*** (10.56)	0.0225*** (11.61)	0.0114*** (7.74)
cold				-0.0175*** (-10.76)	-0.0212*** (-12.45)	-0.0182*** (-9.60)
pitcher_ability	0.720*** (57.18)	4.713*** (68.32)	0.855*** (71.54)	0.825*** (61.18)	0.830*** (62.18)	0.836*** (64.83)
batter_ability	0.827*** (80.34)	5.604*** (102.78)	0.833*** (77.97)	0.829*** (79.94)	0.829*** (80.25)	0.833*** (77.47)
samehand	0.0182*** (10.13)	0.128*** (10.25)	0.0177*** (9.43)	0.0182*** (10.09)	0.0182*** (10.09)	0.0178*** (9.50)
pitcher_home	0.00818*** (9.76)	0.0547*** (9.58)	0.00782*** (8.74)	0.00862*** (9.99)	0.00853*** (9.95)	0.00847*** (9.45)
Observations	1128156	1128156	1007662	1128156	1128156	1037519

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Results: Summary of Magnitude, Normalized

Statistic	Distribution of Ability		Magnitude of Hot Hand Effect		
	Mean	Std Dev	Hot – Cold	Hot-Cold Mean	Hot-Cold Std Dev
Hits Bat	0.268	0.0223	0.0157	5.85%	0.703
On Base Bat	0.339	0.0299	0.0282	8.32%	0.943
Home Runs Bat	0.0304	0.0169	0.0157 [†]	51.71% [†]	0.930 [†]
Strike Outs Bat	0.191	0.059	0.0316	16.54%	0.536
Walks Bat	0.0966	0.0302	0.0333	34.47%	1.103
Hits Pitch	0.264	0.028	0.0153	5.80%	0.547
On Base Pitch	0.335	0.0284	0.0257	7.67%	0.905
Home Runs Pitch	0.0307	0.0098	0.0068 [†]	22.21% [†]	0.692 [†]
Strike Outs Pitch	0.191	0.0538	0.0437	22.88%	0.812
Walks Pitch	0.0965	0.0232	0.0108	11.18%	0.465
Average Bat				23.38%	0.843
Average Pitch				14.02%	0.687
Overall Average				18.70%	0.765

Simulations

We simulate the 3-state Markov model to get some sense of what kind of underlying model would generate our estimates.

- Given measurement error of the state, the underlying streakiness will have to be much larger than our observed outcomes
- Plausibly, managers or players could observe the state better than we can (using more than outcome data), and benefit strategically
- Key parameter: Δ

$$\Pr(Y = 1|hot) = \Pr(Y = 1|avg) + \Delta$$

$$\Pr(Y = 1|cold) = \Pr(Y = 1|avg) - \Delta$$

- Transition probabilities such that steady state distribution is:

$$\Pr(hot) = \Pr(cold) = \kappa$$

for $\kappa = 1\%, 5\%, 10\%$ and $T_{1/2} = 25$.

Simulation Results

- Our regression results are consistent with Δ that is roughly **3-5 times larger** than implied by the empirical estimates.
- For hits:

$$\mu = .270, \Delta = 0.045, \Pr(\text{hot}) = \Pr(\text{cold}) = 5\%$$

generates roughly similar to empirical estimates.

$\implies 2\Delta = 0.09$ is $0.09/0.27 = 33\%$ increase in likelihood of success

- Requires $> 100k$ observations to get (weakly) significant estimates.

Simulation Results

Magnitude of Streakiness

Frequency	$\Delta = 0.01$	$\Delta = 0.02$	$\Delta = 0.04$	$\Delta = 0.05$	$\Delta = 0.10$
$\kappa = 1\%$	0.0105*** (2.98)	0.0120*** (3.40)	0.0139*** (3.94)	0.0158*** (4.48)	0.0289*** (8.24)
$\kappa = 5\%$	0.0116*** (3.28)	0.0149*** (4.22)	0.0252*** (7.19)	0.0328*** (9.39)	0.0924*** (27.36)
$\kappa = 10\%$	0.0122*** (3.44)	0.0184*** (5.23)	0.0389*** (11.18)	0.0543*** (15.73)	0.1606*** (49.63)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Analysis of simulated data for players of heterogenous ability:

$\mu \sim \mathcal{N}(0.270, 0.025)$ and $T_{1/2} = 25$. The first number in each box is the mean estimated γ coefficient across $N = 100$ simulations.

Summary of Hot Hand Evidence

- The difference between being hot and cold is between .5 and 1 standard deviation of ability
 - This is big enough difference to rationalize playing hot players and sitting cold players who are not stars
- Simulations suggest larger magnitude in the true underlying process. We only have a noisy measure of true state.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
- Walks are strongly predicted by both history of past walks and history in home runs and extra base hits.
 - Will use this to test whether teams make correct inference.

Summary of Hot Hand Evidence

- The difference between being hot and cold is between .5 and 1 standard deviation of ability
 - This is big enough difference to rationalize playing hot players and sitting cold players who are not stars
- Simulations suggest larger magnitude in the true underlying process. We only have a noisy measure of true state.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
- Walks are strongly predicted by both history of past walks and history in home runs and extra base hits.
 - Will use this to test whether teams make correct inference.

Summary of Hot Hand Evidence

- The difference between being hot and cold is between .5 and 1 standard deviation of ability
 - This is big enough difference to rationalize playing hot players and sitting cold players who are not stars
- Simulations suggest larger magnitude in the true underlying process. We only have a noisy measure of true state.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
- Walks are strongly predicted by both history of past walks and history in home runs and extra base hits.
 - Will use this to test whether teams make correct inference.

Summary of Hot Hand Evidence

- The difference between being hot and cold is between .5 and 1 standard deviation of ability
 - This is big enough difference to rationalize playing hot players and sitting cold players who are not stars
- Simulations suggest larger magnitude in the true underlying process. We only have a noisy measure of true state.
- Similar results across hitters and pitchers; evidence of both a hot arm and a hot bat.
- Walks are strongly predicted by both history of past walks and history in home runs and extra base hits.
 - Will use this to test whether teams make correct inference.

Results: Endogenous Response

	(1) Walk	(2) Walk	(3) Walk	(4) Walk	(5) Walk
hr_ab_batter_l	0.121*** (11.78)			0.114*** (11.30)	
lagged_hr_control_y	0.150*** (8.41)			0.150*** (8.44)	
hr_ab_batter_ym1	0.374*** (7.64)			0.383*** (7.74)	
hr_ab_batter_ym2	0.280*** (6.96)			0.277*** (6.78)	
exbase_ab_batter_l		0.0733*** (12.15)			0.0736*** (11.52)
lagged_exbase_control_y		0.0831*** (7.46)			0.0885*** (7.42)
exbase_ab_batter_ym1		0.231*** (7.98)			0.275*** (8.09)
exbase_ab_batter_ym2		0.142*** (5.91)			0.143*** (5.58)
hit_ab_batter_l			0.0171*** (3.96)	0.00957* (2.47)	-0.00370 (-0.83)
lagged_hit_control_y			0.00628 (0.88)	-0.000242 (-0.04)	-0.0167* (-2.37)
hit_ab_batter_ym1			0.00205 (0.13)	-0.0108 (-0.74)	-0.0652*** (-3.91)
hit_ab_batter_ym2			0.0267* (2.19)	0.00368 (0.34)	-0.00990 (-0.82)
<i>N</i>	1343450	1343870	1343450	1343450	1343444

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Endogenous responses - Walks

- Walks are predicted by recent performance in home runs and extra-base hits (but not singles)
- Suggestive of defensive response. Pitchers “pitch around” better or hotter hitters, leading to more walks.
- Can use walks to see if teams make correct inference about the magnitude of the hot-hand effect.

Basic Idea: If opposing teams are making correct inferences about the hot hand, they should treat equivalent increases in expected outcome the same.

- A batter who is 20% more likely to hit a HR because he is hot should be treated the same as a batter who is 20% more likely to hit a HR because he is good.

Endogenous responses - Walks

- Walks are predicted by recent performance in home runs and extra-base hits (but not singles)
- Suggestive of defensive response. Pitchers “pitch around” better or hotter hitters, leading to more walks.
- Can use walks to see if teams make correct inference about the magnitude of the hot-hand effect.

Basic Idea: If opposing teams are making correct inferences about the hot hand, they should treat equivalent increases in expected outcome the same.

- A batter who is 20% more likely to hit a HR because he is hot should be treated the same as a batter who is 20% more likely to hit a HR because he is good.

Results: Test for Correct Inference (25 attempts)

	(1) Home run	(2) Walk	(3) Walk/Hr	(4) Extra Base	(5) Walk	(6) Walk/Ex
hr_ab_batter_l	0.0779*** (14.37)	0.121*** (11.78)	1.55			
lagged_hr_control_y	0.111*** (13.49)	0.150*** (8.41)	1.35 (0.99)			
hr_ab_batter_ym1	0.328*** (13.47)	0.374*** (7.64)	1.13 [†] (5.42)			
hr_ab_batter_ym2	0.193*** (10.62)	0.280*** (6.96)	1.45 (0.22)			
exbase_ab_batter_l				0.0575*** (12.27)	0.0733*** (12.15)	1.27
lagged_exbase_control_y				0.0840*** (12.14)	0.0831*** (7.46)	0.99 (2.45)
exbase_ab_batter_ym1				0.270*** (16.31)	0.231*** (7.98)	0.854 ^{††} (7.00)
exbase_ab_batter_ym2				0.134*** (9.80)	0.142*** (5.91)	1.06 (1.12)
<i>N</i>	1211239	1343450		1211233	1343870	

t statistics in parentheses in Columns (1)-(2) and (4)-(5)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[†] reject at 5% level, ^{††} reject at 1% level

Results: Test for Correct Inference (5 and 10 attempts)

	(1) Homerun	(2) Walk	(3) Ratio	(4) Homerun	(5) Walk	(6) Ratio
hr_ab_1.10	0.0313*** (9.14)	0.0736*** (13.38)	2.353			
hr_ab_11.20	0.0300*** (9.33)	0.0325*** (6.03)	1.083††† (14.34)			
hr_ab_21.30	0.0235*** (6.25)	0.0268*** (5.04)	1.142†† (8.36)			
hr_ab_1.5				0.0141*** (5.96)	0.0555*** (14.23)	3.944
hr_ab_6.10				0.0174*** (6.91)	0.0183*** (5.45)	1.054††† (15.48)
hr_ab_11.15				0.0172*** (7.08)	0.0209*** (5.66)	1.215††† (12.99)
hr_ab_16.20				0.0130*** (5.90)	0.0120*** (3.31)	0.920††† (17.27)
hr_ab_21.25				0.0162*** (6.19)	0.0144*** (4.05)	0.884††† (18.13)
lagged_hr_control_y	0.110*** (13.40)	0.149*** (8.39)	1.352†† (8.97)	0.111*** (13.49)	0.150*** (8.43)	1.354††† (14.36)
hr_ab_batter_ym1	0.326*** (13.45)	0.368*** (7.56)	1.129††† (15.34)	0.328*** (13.48)	0.372*** (7.59)	1.133††† (17.14)
hr_ab_batter_ym2	0.192*** (10.63)	0.277*** (6.94)	1.445†† (6.87)	0.193*** (10.62)	0.279*** (6.96)	1.446††† (12.83)
<i>N</i>	1211239	1343876		1211239	1343876	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

† reject at 5% level, †† reject at 1% level, ††† reject at 0.1% level

Main Findings

- Both hotter and better hitters in home runs and extra base hits are walked more frequently.
- Teams appear to make correct inferences about the magnitude of the hot hand effect based on prior 25 attempts.
 - Walks increase with hot hitters in roughly the same manner as better hitters.
- However, there is a tendency to overreact to very recent outcomes.
 - Walks increase more than optimal inferences would indicate.
 - Consistent with a version of the hot hand fallacy
 - * Albeit somewhat weaker and more limited than what one usually hears.

Summary of Endogenous Responses and Inference

Main Findings

- Both hotter and better hitters in home runs and extra base hits are walked more frequently.
- Teams appear to make correct inferences about the magnitude of the hot hand effect based on prior 25 attempts.
 - Walks increase with hot hitters in roughly the same manner as better hitters.
- However, there is a tendency to overreact to very recent outcomes.
 - Walks increase more than optimal inferences would indicate.
 - Consistent with a version of the hot hand fallacy
 - * Albeit somewhat weaker and more limited than what one usually hears.

Cognitive Mistakes or Equilibrium Adjustments?

We argue that many were too quick in labeling the lack of a hot hand as a pervasive cognitive mistake.

- To us, a more natural interpretation, with basketball results, is that this was simply efficient equilibrium behavior
- Given endogenous responses, equilibrium outcomes will look different from partial equilibrium
- Easy to confuse equilibrium with behavioral effect
- Similarly, it is easy to confuse “no effect” with an equilibrium response.

Conclusion

- Formulated a general hypothesis and simple test as to where one should expect to find evidence of a hot hand.
 - Baseball vs basketball
- We find strong evidence for a hot hand in all ten baseball statistics that we analyze
 - The effect is significant, consistent with strategic decisions
 - Indicative of a stronger underlying process
- We also document evidence that teams respond to streaky behavior in a manner roughly consistent with making correct inferences.
 - Except for a tendency to overreact to very recent performance
- Future work: explore the mechanism underlying this finding.