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

Brett Green – UC Berkeley
Jeffrey Zwiebel – Stanford University

April 2015

# Motivation

- 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).
  - Generalizes to settings outside of sports
  - 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 | P(hit/3 misses) | P(hit/2 misses) | P(hit/1 miss) | P(hit) | P(hit/1 hit) | P(hit/2 hits) | P(hit/3 hits) | Serial correlation r |
|---|---|---|---|---|---|---|---|---|
| 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

# Drawbacks of this Literature

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

# 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.

# 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 paper

- 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.

# Related Literature

- Lots of papers that do not find a hot-hand effect.
- According to a survey (Bar-Eli, 2006): *The empirical evidence for the existence of the hot hand is considerably limited*.

Evidence in Favor of the Hot Hand

- Horseshoe Pitching: Smith (2003)
- Bowling: Dorsey-Palmateer and Smith (2004)
- Free throws: Arkes (2010), Miller and Sanjurjo (2014)

Endogenous Responses to Streaky Behavior

- Basketball: Aharoni and Sarig (2012), Bocskocsky et al. (2014)

# 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

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

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

**In equilibrium**

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

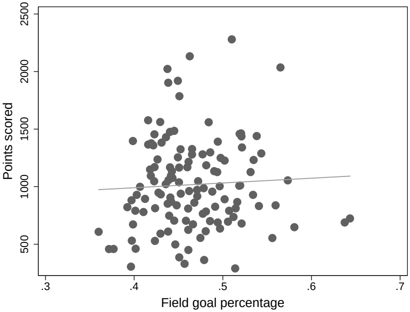
# Endogenous responses: Basketball vs Baseball

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

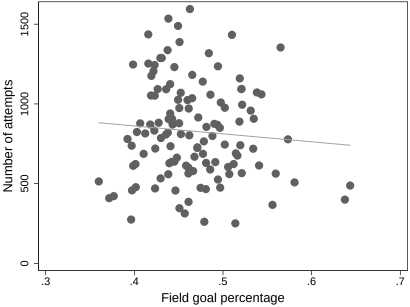
- Defense should optimize batter by batter
- Little scope for transferring resources across batters
- Also, no reason such transfers would be optimal anyway
  - Batter order is fixed: no analogous offensive adjustment in baseball.
- One significant (unobservable) adjustment: "pitch around"
  - Doing 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

# Basketball



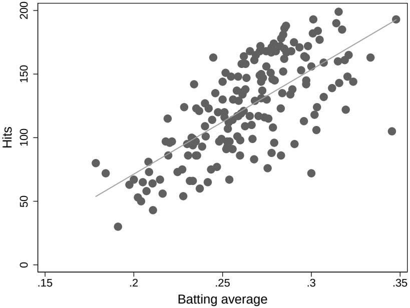

(a) Successes



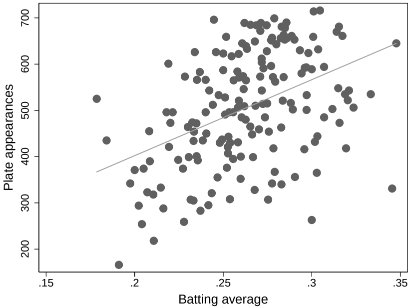

(b) Attempts

- 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.

# 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.

# 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:

$$Hot_{it} = 1\left(R_{it}^{L} \geq Ability_{it} + A\right)$$
$$Cold_{it} = 1\left(R_{it}^{L} \leq Ability_{it} - B\right)$$

   - Proportional Cutoffs:

$$Hot_{it} = 1\left(R_{it}^{L} \geq q \cdot Ability_{it}\right)$$
$$Cold_{it} = 1\left(R_{it}^{L} \leq r \cdot Ability_{it}\right)$$

# Measuring state

3) Distribution-based state measure:

$$State^d_{it} = freq\{R^L_{it} > R^L_{is}, \forall s \notin \mathcal{S}(t)\}$$

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:

$$Hot_{it} = 1\left(State^d_{it} > 1 - p\right)$$
$$Cold_{it} = 1\left(State^d_{it} < p\right)$$

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 |

# Overview of Hot Hand Evidence

- Positive and strongly significant $\gamma$ for all categories in virtually all of our specifications and tests:
  - A .400 increase in OBP in the last 25 PA $\Longrightarrow$ 26 extra OBP points
  - A .400 increase in BA in the last 25 AB $\Longrightarrow$ 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*** | 0.292*** | 0.0114*** | | | |
| | (14.61) | (14.76) | (8.63) | | | |
| hot | | | | 0.0164*** | 0.0169*** | 0.00762*** |
| | | | | (7.95) | (8.61) | (4.45) |
| cold | | | | -0.0133*** | -0.0134*** | -0.00958*** |
| | | | | (-7.35) | (-7.43) | (-4.81) |
| batter_ability | 0.551*** | 2.440*** | 0.615*** | 0.602*** | 0.602*** | 0.603*** |
| | (20.92) | (22.36) | (20.37) | (21.30) | (21.31) | (20.01) |
| pitcher_ability | 0.578*** | 2.597*** | 0.576*** | 0.580*** | 0.580*** | 0.577*** |
| | (44.25) | (44.18) | (42.86) | (44.34) | (44.31) | (43.12) |
| samehand | -0.0245*** | -0.110*** | -0.0247*** | -0.0246*** | -0.0246*** | -0.0249*** |
| | (-19.49) | (-19.60) | (-18.94) | (-19.32) | (-19.33) | (-18.94) |
| batter_home | 0.0127*** | 0.0569*** | 0.0129*** | 0.0128*** | 0.0128*** | 0.0129*** |
| | (16.08) | (16.13) | (15.82) | (16.12) | (16.12) | (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*** | 1.945*** | 0.00911*** | | | |
| | (15.86) | (15.83) | (10.93) | | | |
| hot | | | | 0.00420*** | 0.00786*** | 0.00190*** |
| | | | | (6.03) | (9.18) | (3.79) |
| batter_ability | 0.702*** | 19.13*** | 0.814*** | 0.769*** | 0.752*** | 0.765*** |
| | (51.22) | (37.55) | (58.33) | (55.89) | (54.41) | (55.11) |
| pitcher_ability | 0.376*** | 11.32*** | 0.378*** | 0.378*** | 0.377*** | 0.378*** |
| | (20.49) | (22.72) | (19.66) | (20.49) | (20.50) | (20.50) |
| samehand | -0.00452*** | -0.124*** | -0.00454*** | -0.00442*** | -0.00446*** | -0.00442*** |
| | (-8.92) | (-7.65) | (-8.59) | (-8.58) | (-8.69) | (-8.58) |
| batter_home | 0.00209*** | 0.0562*** | 0.00196*** | 0.00207*** | 0.00208*** | 0.00206*** |
| | (5.35) | (4.84) | (4.81) | (5.25) | (5.29) | (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*** | 0.155*** | 0.00534*** | | | |
| | (5.98) | (6.00) | (3.44) | | | |
| hot | | | | 0.00787*** | 0.00644*** | 0.00427* |
| | | | | (4.04) | (3.66) | (2.38) |
| cold | | | | -0.00728*** | -0.00921*** | -0.00447* |
| | | | | (-3.41) | (-4.22) | (-2.07) |
| batter_ability | 0.366*** | 1.839*** | 0.386*** | 0.388*** | 0.389*** | 0.382*** |
| | (15.94) | (16.17) | (14.67) | (16.22) | (16.32) | (14.81) |
| pitcher_ability | 0.529*** | 2.688*** | 0.528*** | 0.530*** | 0.530*** | 0.530*** |
| | (33.80) | (33.15) | (32.64) | (33.80) | (33.80) | (32.82) |
| samehand | -0.0131*** | -0.0657*** | -0.0130*** | -0.0130*** | -0.0131*** | -0.0131*** |
| | (-12.09) | (-12.05) | (-11.68) | (-12.03) | (-12.03) | (-11.83) |
| batter_home | 0.00911*** | 0.0458*** | 0.00899*** | 0.00913*** | 0.00913*** | 0.00897*** |
| | (10.84) | (10.87) | (10.32) | (10.85) | (10.84) | (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*** | 0.766*** | 0.0288*** | | | |
| | (21.99) | (22.91) | (17.53) | | | |
| hot | | | | 0.0195*** | 0.0225*** | 0.0114*** |
| | | | | (10.56) | (11.61) | (7.74) |
| cold | | | | -0.0175*** | -0.0212*** | -0.0182*** |
| | | | | (-10.76) | (-12.45) | (-9.60) |
| pitcher_ability | 0.720*** | 4.713*** | 0.855*** | 0.825*** | 0.830*** | 0.836*** |
| | (57.18) | (68.32) | (71.54) | (61.18) | (62.18) | (64.83) |
| batter_ability | 0.827*** | 5.604*** | 0.833*** | 0.829*** | 0.829*** | 0.833*** |
| | (80.34) | (102.78) | (77.97) | (79.94) | (80.25) | (77.47) |
| samehand | 0.0182*** | 0.128*** | 0.0177*** | 0.0182*** | 0.0182*** | 0.0178*** |
| | (10.13) | (10.25) | (9.43) | (10.09) | (10.09) | (9.50) |
| pitcher_home | 0.00818*** | 0.0547*** | 0.00782*** | 0.00862*** | 0.00853*** | 0.00847*** |
| | (9.76) | (9.58) | (8.74) | (9.99) | (9.95) | (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

| | Distribution of Ability | | Magnitude of Hot Hand Effect | | |
|---|---|---|---|---|---|
| **Statistic** | **Mean** | **Std Dev** | **Hot − Cold** | $\frac{\text{Hot-Cold}}{\text{Mean}}$ | $\frac{\text{Hot-Cold}}{\text{Std Dev}}$ |
| 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: $\Delta$

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

- 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 $\Delta$ that is roughly **3-5 times larger** than implied by the empirical estimates.
- For hits:

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

generates roughly similar to empirical estimates.

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

- Requires $> 100$k 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 $\gamma$ 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.

# Results: Endogenous Response

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

# Results: Test for Correct Inference (25 attempts)

| | (1)<br>Home run | (2)<br>Walk | (3)<br>Walk/Hr | (4)<br>Extra Base | (5)<br>Walk | (6)<br>Walk/Ex |
|---|---|---|---|---|---|---|
| hr_ab_1_25 | 0.0779*** | 0.121*** | 1.55 | | | |
| | (14.37) | (11.78) | | | | |
| hr_ab_batter_lagged | 0.111*** | 0.150*** | 1.35 | | | |
| | (13.49) | (8.41) | (0.99) | | | |
| hr_ab_batter_ym1 | 0.328*** | 0.374*** | $1.13^{\dagger}$ | | | |
| | (13.47) | (7.64) | (5.42) | | | |
| hr_ab_batter_ym2 | 0.193*** | 0.280*** | 1.45 | | | |
| | (10.62) | (6.96) | (0.22) | | | |
| exbase_ab_1_25 | | | | 0.0575*** | 0.0733*** | 1.27 |
| | | | | (12.27) | (12.15) | |
| exbase_ab_batter_lagged | | | | 0.0840*** | 0.0831*** | 0.99 |
| | | | | (12.14) | (7.46) | (2.45) |
| exbase_ab_batter_ym1 | | | | 0.270*** | 0.231*** | $0.854^{\dagger\dagger}$ |
| | | | | (16.31) | (7.98) | (7.00) |
| exbase_ab_batter_ym2 | | | | 0.134*** | 0.142*** | 1.06 |
| | | | | (9.80) | (5.91) | (1.12) |
| $N$ | 1211239 | 1343450 | | 1211233 | 1343870 | |

$^{\dagger}$ reject at 5% level, $^{\dagger\dagger}$ 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*** | 0.0736*** | 2.353 | | | |
| | (9.14) | (13.38) | | | | |
| hr_ab_11_20 | 0.0300*** | 0.0325*** | 1.083††† | | | |
| | (9.33) | (6.03) | (14.34) | | | |
| hr_ab_21_30 | 0.0235*** | 0.0268*** | 1.142†† | | | |
| | (6.25) | (5.04) | (8.36) | | | |
| hr_ab_1_5 | | | | 0.0141*** | 0.0555*** | 3.944 |
| | | | | (5.96) | (14.23) | |
| hr_ab_6_10 | | | | 0.0174*** | 0.0183*** | 1.054††† |
| | | | | (6.91) | (5.45) | (15.48) |
| hr_ab_11_15 | | | | 0.0172*** | 0.0209*** | 1.215††† |
| | | | | (7.08) | (5.66) | (12.99) |
| hr_ab_16_20 | | | | 0.0130*** | 0.0120*** | 0.920††† |
| | | | | (5.90) | (3.31) | (17.27) |
| hr_ab_21_25 | | | | 0.0162*** | 0.0144*** | 0.884††† |
| | | | | (6.19) | (4.05) | (18.13) |
| hr_ab_batter_lagged | 0.110*** | 0.149*** | 1.352†† | 0.111*** | 0.150*** | 1.354††† |
| | (13.40) | (8.39) | (8.97) | (13.49) | (8.43) | (14.36) |
| hr_ab_batter_ym1 | 0.326*** | 0.368*** | 1.129††† | 0.328*** | 0.372*** | 1.133††† |
| | (13.45) | (7.56) | (15.34) | (13.48) | (7.59) | (17.14) |
| hr_ab_batter_ym2 | 0.192*** | 0.277*** | 1.445†† | 0.193*** | 0.279*** | 1.446††† |
| | (10.63) | (6.94) | (6.87) | (10.62) | (6.96) | (12.83) |
| *N* | 1211239 | 1343876 | | 1211239 | 1343876 | |

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

# 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

Many other examples in the literature:

- Active fund manager ability and fund flows (Berk and Green)
- Overconfidence vs. Agency/Incentives
- (In)attention vs. Rational Information Processing

# What is Behavioral?

Consider the old debate between efficient market proponents and behavioralists in finance:

- EM proponents argued that the market followed a random walk
- Widespread beliefs to the contrary are due to cognitive mistakes perceiving patterns in random data
- Behavioralists argued that not only is there predictability, but that we should expect this due to many behavioral effects of investors (joint with limits to arbitrage)

# What is Behavioral?

Now consider the hot hand debate:

- Behavioralist argue that sports players outcomes are random
- Widespread beliefs to the contrary are due to cognitive mistakes perceiving patterns in random data
- We argue that not only are there patterns, but that we should expect this due to many behavioral (and physical) effects on players

Is our conclusion behavioral?

# 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.

# Learning or Streakiness?

Our hot hand evidence can be interpreted in two ways:

1) Streakiness: ability changes over time at a high frequency.
2) Learning: ability does not change over time but recent performance provides additional information about long-run ability.

To distinguish between these two, we used simulations and

- Compare our coefficients to the predicted value under learning only.
- Change our definition of long-run ability to include the most recent 50 outcomes.
- Lag the state variable within the window.

All three additional tests yield similar findings:

- Each interpretation accounts for about 50% of the hot-hand effect.