

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

Brett Green* Jeffrey Zwiebel†

February 19, 2016

Abstract

We test for a “hot hand” in Major League Baseball using panel data. We find strong evidence for its existence in all ten statistical categories we consider. The magnitudes are significant; being “hot” corresponds to between one-half and one standard deviation in the distribution of player abilities. Our results are in notable contrast to the majority of the hot-hand literature, which has found little to no evidence for a hot hand in sports, often employing basketball shooting data. We argue that this difference is attributable to endogenous defensive responses: basketball presents sufficient opportunity for transferring defensive resources to equate shooting probabilities across players whereas baseball does not. We then document that baseball teams do respond to recent success in their opponents’ batting performance. Our results suggest that teams use recent performance in a manner consistent with drawing a correct inference about the magnitude of the hot-hand, but have a tendency to overreact to very recent performance (i.e., the last 5 attempts).

1 Introduction

The absence of a “hot hand” in sports is frequently cited in behavioral economics as an example of a widespread cognitive mistake. Following Gilovich et al. (1985)’s (GVT) seminal study of streakiness in basketball shooters, a large literature has examined the hot hand. Most of this literature follows GVT in purporting to show that sports players do not have a propensity for hot or cold streaks, contrary to the almost universally held perceptions among players, coaches, and fans alike. That is, temporal outcomes are not found to exhibit streaks beyond what would follow randomly under

*Haas School of Business, University of California, Berkeley, CA 94720, email: greenb@berkeley.edu, webpage: <http://faculty.haas.berkeley.edu/bgreen/>

†Graduate School of Business, Stanford University, Stanford, CA 94305, email: zwiebel@stanford.edu

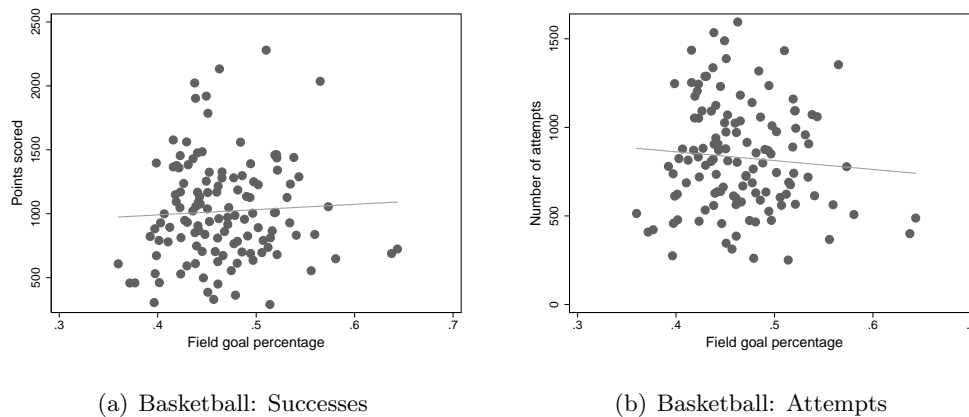


Figure 1 – Relation between points scored, attempts and success percentage in the National Basketball Association (NBA) using data from the 2012-13 season. Figures include data from all players who started in at least 50 percent of all regular season games. In neither category is the correlation coefficient statistically different from zero.

independence, and consequently, widespread perceptions of hot hands are attributed to a basic cognitive mistake of inferring patterns in random data.¹

Much of this hot-hand literature, however, is plagued by two critical drawbacks. First, it often fails to take endogenous strategic responses into account. A hot shooter in basketball should be defended more intensely, and should take more difficult shots, both of which will lower his shooting percentage. And second, many of the tests conducted in this literature, especially tests attempting to address this endogeneity, lack the power to identify plausible and significant models of hot hands.

To get an intuitive sense of the importance of endogenous responses in sports, consider field goal shooting in basketball, the focus of GVT and a large amount of the subsequent literature on the hot hand. One might be inclined to think that better players make a higher percentage of the field goals they attempt. However, this is generally not the case. Consider shooting statistics from the 2012-13 season for the NBA. Not surprisingly, the leading scorers in the league are typically acknowledged to be the best offensive players in the game: For the top 20 scorers (given by points scored per game), all 20 are considered “star” players, 14 of them were selected for the all-star game, and almost all are among the most highly paid players in the game. Yet, the median (mean) field goal percentage among these top 20 scorers was 0.447 (.465), which is slightly *below*, but statistically indistinguishable from, that of a *typical* starting player .460 (.466).² Figure 1(a) depicts the relationship (or lack thereof) between number of points scored and the likelihood of success.

There is an obvious and widely accepted explanation for why the NBA scoring leaders and

¹GVT (p. 311) states: “Evidently, people tend to perceive chance shooting as streak shooting and they expect sequences exemplifying chance shooting to contain many more alternations than would actually be produced by a random (chance) process. Thus, people ‘see’ a positive serial correlation in independent sequences, and they fail to detect a negative serial correlation in alternating sequences.”

²A typical starting player is defined as one who started in at least 50 percent of all regular season games.

universally acknowledged stars do not collectively shoot any better than the league average. Even a very casual basketball observer can readily observe that Kobe Bryant (career shooting percentage of .454) draws far more defensive attention, and attempts far more difficult shots, than his teammates. Indeed, a fundamental aspect of defensive strategy in basketball involves allocating more defensive attention to better shooters and less to weaker shooters. Insofar as defenses have great flexibility in allocating their defensive attention to cover the five opposing players on the court, and offenses are free in turn to allocate their shots across their players on the court, it would be natural to expect, to a first approximation, that such adjustments would yield shooting percentages that are equated across players on a given team. If Kobe Bryant's probability of making a shot was higher than that of teammate Jordan Hill, defenses should cover Bryant more and Hill less, and Bryant should shoot more and Hill less, until these probabilities were equated.³ Hence, in equilibrium, individual shooting percentages are not likely to be indicative of players' shooting abilities; good shooters are likely to shoot more, and raise the overall team shooting percentage, but should not be expected to have a higher shooting percentage than their teammates. This also implies there should not be a relationship between the shooting percentage and the number of attempts, which is consistent with Figure 1(b).⁴

Few would dispute the importance of defensive adjustments when evaluating field goal percentages across players. No one that we are aware of argues that Kobe Bryant is but an average shooter, and no one disputes that he is defended more and attempts harder shots than others. The same reasoning should also apply if a player is "hot". That is, if better shooters are defended to the point where their shooting percentage reverts to average, one should expect that this would also hold for players who are *temporarily* better than average, i.e., for hot players. A hot player, who is temporarily shooting better than he normally does, would temporarily exceed his teammates' shooting percentage if he only took his usual shots and received his usual defensive attention. In response, he should attempt more difficult shots, and he should receive more defensive attention, until these adjustments lower his marginal shooting percentage to equate it with his that of his teammates.⁵ Consequently, for the same reason good shooters are unlikely to exhibit a higher shooter percentage than average players, "hot" shooters are unlikely to exhibit a higher future shooting percentage than they typically realize. As such, it is rather surprising that tests finding no evidence that recent shooting success in basketball predicts future successes have been interpreted by some as

³There are some qualifications and limitations to this simple argument, based on specific features of the game – i.e., 3-point shots are worth more, shots by inside players are harder to set-up than outside shots, etc. See footnote 11 for further elaboration on this point.

⁴Aharoni and Sarig (2012) also provide evidence that is consistent with these arguments: good shooters and hot shooters both shoot more and raise overall team shooting percentages.

⁵Huizinga and Weil (2009) have found that shooters who have recently made several shots in a row shoot more frequently. They interpret this as evidence that defensive adjustments are limited. This interpretation, however, does not seem to follow from the result, as one might expect both offensive adjustments (more shots) and defensive adjustments (more defensive attention) in response to a better shooter. Indeed, the Huizinga-Weil interpretation here of limited defensive responses would be akin to interpreting Kobe Bryant's high shooting frequency as evidence that he does not receive extra defensive attention, which seems implausible.

compelling evidence of a cognitive mistake by players, coaches, and fans (i.e., a hot-hand fallacy). It seems a more natural interpretation a priori would be one of strategic equilibrium adjustments by offenses and defenses in the face of short-term changes in ability.

Also problematic is the lack of power in many of the tests that have been employed in research on the hot hand. This is especially notable in attempts to address the issue of endogeneity. For example, GVT also consider free-throws in Study 4 of their paper. Since free-throws are not defended, and are taken from the same spot of the floor, they do not suffer from the endogeneity of field goals. GVT look at whether the likelihood of success of a second free-throw is related to the success of a first free-throw. However, they only consider nine players, with no more than 430 pairs of free-throws for the player who shot the most, and less than 200 pair of free-throws for five of the nine players. Along similar lines, many of the tests by others designed to address endogeneity have looked for evidence of a hot hand player by player, often relying on a few hundred, or at most, a few thousand observations per player.⁶ Under simple specifications for a hot hand, that seem well in line with common perceptions on the magnitude of streakiness, one needs a dataset at least an order of magnitude larger than this in order to be able to distinguish randomness from a hot hand and even then most tests are biased against finding evidence of streakiness.⁷

In this paper we look for evidence of streakiness in major league baseball. Baseball data is particularly well-suited to address both the drawback of lack of power in the tests and the drawback of endogenous defensive responses, and moreover, affords us a manner to quantify an aspect of endogenous defensive responses and to ask whether these responses are justified.

On the first point, statistical power, our tests have high statistical power, for several reasons. First, there is an abundance of baseball data. There are roughly 200,000 batter plate appearances per year (the unit of our observation), and we employ twelve years of data, giving us over two million observations. Second, we look for general evidence of streakiness over the panel of our data instead of considering each player individually. Given this, our tests have a high degree of power to discern reasonable models of streakiness from random outcomes.

More interesting conceptually, we will argue in Section 2 that the scope for transferring defensive resources across players is far more limited in baseball than in basketball, and insufficient to equate margins across players as in basketball. In brief, the argument is that while in basketball the offensive team chooses on each possession who will shoot the ball and the defensive team can allocate defensive resources to minimize the value of the best shot available, in contrast, in baseball all batters in the game face a pitcher individually and sequentially, according to a predetermined order. Since each batter is faced one at a time, and since the same defensive resources are available for each batter, there is relatively little scope for the defensive team to shift resources from defending one batter to another. Consequently, whereas optimal equilibrium responses in basketball should

⁶Several recent exceptions include Arkes (2010), Miller and Sanjurjo (2014) and Bocskocsky et al. (2014). See Section 2.2 for more discussion of related literature.

⁷Related points appear in Stern and Morris (1993), Wardrop (1999), Miyoshi (2000), Dorsey-Palmateer and Smith (2004), Stone (2012), and Arkes (2013).

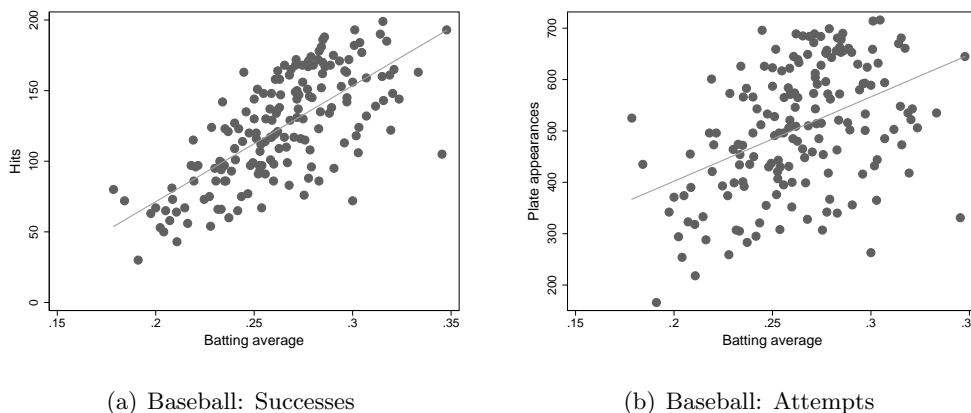


Figure 2 – Relation between number successes, number of attempts and success percentage in Major League Baseball using data from the 2013 season. Figures include data from all players who started in at least 50 percent of all regular season games. In both relationships, the coefficient is positive and statistically significant.

eliminate evidence of streakiness in shooting data, one should still expect to see such streakiness (if it exists) in baseball hitting and pitching data. This is suggested by Figure 2, which indicates that good players in baseball are both more likely to be successful and have more attempts. Notice that this lies in stark contrast with analogous plots for basketball in Figure 1.

There is one significant adjustment that can be made which partially transfers defensive resources across hitters: A pitcher can “pitch around” a good hitter or a hot hitter, by not giving the hitter a good pitch to hit. As we discuss in Section 2, this activity will not directly affect our hitting statistics, which net out walks from plate appearances. Consequently, walks notwithstanding, if a hot hand exists among baseball players, recent past performance should predict future outcomes for our statistics. Moreover, we can measure the degree to which a defense responds by pitching around hot hitters by measuring how walks vary with recent performance. It is well known that better hitters walk more frequently, due in part to being pitched around more frequently. We test whether hot hitters are also walked more, and then use the relationship between hitter ability and walks to back out what teams seem to be inferring about the magnitude and longevity of streakiness. We then compare that inference with our own estimates on streakiness to see whether defenses are reacting appropriately to hot (or cold) hitters.

In our tests, we consider ten different statistical categories in baseball. In particular, we consider five statistics for hitters: batting average, home runs per at bat, on-base percentage, strike outs per at bat, and walks per plate appearance; and the corresponding five statistics for pitchers: batting average allowed, home runs allowed per at bat etc. The first three statistics are “good” outcomes for hitters and “bad” outcomes for pitchers. The fourth statistic is bad for hitters and good for pitchers. With all these statistics we hypothesize that recent performance for a given statistic predicts future performance in the same statistic, for both hitters and pitchers. The fifth statistic,

walks per plate appearance, begets an additional interpretation. Like the first three statistics, a walk is a good outcome for a hitter, and under a hot hand, one would expect past recent walks to predict future walks. However, unlike the other statistics, walks are likely to increase if a pitcher “pitches around” (i.e., tries to avoid pitching directly to) a given hitter. Consequently, this is where we expect to find direct evidence of an endogenous reaction - good hitters and hot hitters should receive fewer good pitches to hit, resulting in more walks.

For each of these statistics we examine whether recent performance is predictive of current outcomes, for both hitters and pitchers. For all our tests, we control for the ability of the hitter and pitcher, as well as situational variables such as the ballpark and platoon differentials (hitters hit worse and pitchers pitch better when facing a same-handed opponent).

Strikingly, we find recent performance is highly significant in predicting performance in all ten statistical categories that we examine (five for batters and five for pitchers). In all cases, being “hot” in a statistic makes one more likely to perform well in the same statistic. Furthermore these effects are of a significant magnitude: for instance, a hot hitter in on-base percentage will exhibit an on-base percentage roughly .028 higher (28 OBP points higher) than when cold after controlling for all other explanatory variables. This compares with an average on-base percentage across players of .339 with a standard deviation of .030. Thus, the difference between a hot batter and a cold batter is roughly equivalent to the difference between a 70th percentile batter and a 50th percentile one. Our results have a similar degree of significance across other statistics. For example, a batter who is “hot” in home-runs is 15-25% more likely (0.5-0.75 percentage points or about one half a standard deviation) more likely to hit a home run in his next at bat.

Consistent with an endogenous defensive response, we also find strong evidence that when a batter has recently hit a relatively high percentage of home runs or extra base hits, he is significantly more likely to draw a walk. Thus, not only do better hitters walk more often, but hotter hitters walk more often as well. In both cases, this is most readily interpreted to being “pitched around”; i.e., greater care is taken when pitching to better and hotter hitters. We then test whether the magnitude of the endogenous response to hot hitters is consistent with teams drawing the correct inference regarding about the magnitude of the hot hand effect. Based on the batters performance in the last 25 at bats, we find that teams seem to respond in a way that is consistent with them drawing correct inferences about the magnitude of the streakiness effect. That is, a median batter not only performs like a 70th percentile batter when hot, he is also walked roughly as frequently as a 70th percentile batter.

As a more refined test, we decompose players recent performance into shorter intervals of length 5 or 10 at bats. We show that players with strong performance in their most recent 5 attempts are disproportionately more likely to be walked (relative to our inference of how much better such a hot player is likely to perform). That is, defenses seem to pitch around (and as a result walk) batters who have hit home runs very recently more frequently than can be justified by the increased likelihood of them hitting a home run in their next attempt. One interpretation of this finding is

consistent with a weak version of a hot hand fallacy: teams overestimate the effect of the last 5 at bats of an opposing hitter on his performance in the next at bat, even while they seem to infer the effect of the last 25 at bats accurately.⁸ On the other hand, these findings are not consistent with the Gambler's Fallacy (Tversky and Kahneman, 1971). If agents incorrectly believe in systematic reversals then one would expect them to *underreact* to recent streaky performance (Rabin and Vayanos, 2010).

It is important to note that given noisy outcomes, we only have a noisy estimate of whether a player is hot or cold from observing his recent outcomes. To assess the true magnitude of streakiness in the underlying data generating process, one needs a model for streakiness. While we are agnostic as to the specific model underlying streaks, for intuition we consider a basic three state Markov model, where players can switch between a normal state, a hot state, and a cold state.⁹ Note that in such a model, the probability of a successful outcome, conditional on a good recent history, will be less than the probability of a success in the hot state, for two reasons: first, the possibility that recent successes were due to luck rather than being in the hot state, and second, the possibility of a switch out of the hot state.¹⁰ Therefore, under this model of streakiness, our regression estimates will tend to understate the magnitude of the effect. The streakiness exhibited in our data corresponds to a model where the difference between hot and cold states is highly significant and 5 times larger than suggested by our empirical results. For instance, our empirical estimates are consistent with a model in which a batter in the hot state is 33 percent more likely to get a hit than a batter in the cold state. This translates to approximately a 80 batting average point (or 8 percentage point) difference between a hot and a cold batter, which also corresponds to a 2.5-3 standard deviation difference.

Our finding, that recent past performance strongly predicts future outcomes, after controlling for player ability (excluding recent performance), can be attributed to streakiness or to learning new information about a player's long-term ability through this short-term performance. In line with the literature, we refer to a "hot hand" as the ability to predict future outcomes from recent past performances, without making the distinction between streakiness and learning. Indeed, this is the strategically relevant definition; how should one expect performance to change conditioning on the recent past. This distinction, however, is interesting, and we undertake several tests in Section 4.2.1 to provide a decomposition of the hot hand into these two components. Our results vary across the different statistics, but in general we find that streakiness accounts for just over

⁸These two findings combined seem to imply that teams underreact to batters' performance that took place more than 5 but less than 25 at bats prior to the current plate appearance. In the data, we find this underreaction to be small and not statistically identifiable.

⁹Two alternative models that strike us as plausible are a setting with a continuous state variable for "hotness", and a setting where switches between states are determined by outcomes rather than exogenous changes in the state. We do not pursue such models here.

¹⁰The first point is similar to one made in Stone (2012), which formally demonstrates a downward bias in the first auto correlation in an AR1 model of streakiness. Arkes (2013) quantifies the downward bias using simulations in a two-state model.

50% of the hot-hand effect, with learning attributable to a bit less than 50% of the effect.

The distinction we make between basketball and baseball motivates a broader question on the existence of a hot hand in other sports and activities. We hypothesize that skilled activity will generally demonstrate both persistent long-run variation in ability and transitory short-run variation in ability – i.e., hot hands. We remain agnostic about the relative importance of different factors in short-run variations in ability, but find it likely that both physical and psychological factors play important roles. We further hypothesize that in certain sports endogenous responses equate margins and thereby confound empirical evidence of a hot hand (i.e. basketball, as well as other team sports such as hockey and soccer), while other sports do not have equating endogenous responses, and will therefore demonstrate streaky short-run outcomes (such as baseball, golf, and bowling). In some cases this distinction should be obvious (e.g., there is no defense in golf or bowling), where in other cases it is more subtle and depends on the details of the sport (e.g., the distinction between basketball and baseball).

In the next section, we discuss the difference between endogenous responses available in basketball and baseball in more detail, in order to highlight the difference in a setting where endogenous responses are likely to equate margins across players and a setting where such responses are not likely to do so. We follow this discussion with a brief summary of the literature. Section 3 discusses the data and our methodology. In Section 4, we present evidence of streakiness in baseball and interpret these results. In Section 5, we present evidence of endogenous defensive responses and test whether such responses are consistent with a hot-hand fallacy. In Section 6, we present a simple model of streakiness to give an intuitive sense of the magnitudes consistent with our empirical findings. Section 7 provides general discussion of the hot hand more broadly in other activities. Section 8 concludes. Most of the empirical results are located in tables in the Appendix. Additional results and robustness checks are relegated to an online appendix.

2 Background

2.1 Basketball vs. Baseball

It is instructive to discuss in detail the different endogenous responses available in basketball and baseball, in order to distinguish between a setting where margins are likely to be equated across players and those where they are not. We first discuss basketball, to provide our argument that endogenous responses are likely to be sufficient to equate margins there, and then turn our attention to baseball, where we will argue that endogenous responses are far more limited.

In basketball all active players are on the court at the same time, and defenses are highly fluid. On each possession, the offensive team passes the ball between players looking for the best shot opportunity, and the defense allocates their defensive resources across the offensive players. Offenses and defenses are dynamic, in constant motion, with nearly continuous adjustments made to obtain and defend against shooting opportunities. These adjustments are frequent, detailed,

and highly-contingent. For example, switches, double-teams, help, posts, and other manners of temporary defensive assistance are given and withdrawn across players fluidly, in manners small and large, often changing in fractions of a second.

A natural model to characterize this strategy is that offenses try to maximize the value of a shot across players and shooting opportunities, and defenses try to minmax across this offensive selection. This should lead to shooting values being equated across offensive players at any given time. That is, if there are better marginal shots available for player A than player B, player A should be shooting more and player B should be shooting less, and defenses should be covering player A more and player B less, until this discrepancy vanishes. Good players, or hot players, should not have better marginal shots than bad players or cold players; both defensive and offensive adjustments should serve to equate such margins.¹¹

Turning to baseball, we presently argue that the ability to transfer resources across players in baseball is far more limited than basketball, and will be insufficient to equate margins across batters.¹² There are several reasons for this difference.

Most importantly, in baseball pitchers face batters sequentially rather than simultaneously. Unlike in basketball, there is not a fixed defensive resource (the number and defensive ability of players on the field) that must be allocated and adjusted across all the batters (offensive players). Rather, each hitter faces the pitcher and his defensive team (his fielders) one at a time, and the pitcher and all fielders can focus almost all of their defensive attention on the hitter at hand. Consequently, if batter A is a better hitter than batter B, the opposing team has little scope to transfer defensive resources from B to A.

Moreover, even in the unlikely event that a team could transfer defensive resources across batters in a sufficient manner, there would be no general reason why it would be optimal for defenses to equate margins across offensive players. Indeed one would want to expend defensive resources on whichever players (or game situations) yielded the highest defensive benefit, and there is little reason to suspect that would typically involve transferring defensive resources from weaker players to stronger ones until averages were equated. Note the contrast here with basketball, where the offensive team chooses *who* will shoot on each possession, and should allocate the shot to whomever has the highest value shot. It is this choice that should in turn lead defenses to minmax across

¹¹It is worth noting that for a number of reasons, the highest value shot for an offense will not be the same as the highest percentage shot. To correctly account for a shot value, one will need to take into account a number of additional factors such as: 3-point shots (i.e., long shots are worth 3 points, while shorter shots are worth 2 points); the likelihood that the offensive team gets the rebound (and hence another shot opportunity) if the shot is missed; the likelihood that the shooter gets fouled taking the shot (and hence gets additional scoring opportunities); and the difficulty in setting up the shot (i.e., the likelihood that the opponent steals the ball prior to the shot). There will also be some inframarginal shot opportunities such as fast breaks. For these and other related reasons, it is not quite accurate to say that marginal shooting percentages should be equated across offensive players. However, as we have argued in prior versions of this paper, these deviations between shot value and shooting percentage are likely to be relatively small, and consequently, a notion of equal marginal shooting percentages is likely to be a good approximation of optimal strategies. For more detail, see an earlier draft of this paper, available from the authors on request.

¹²We speak of batters here; the argument for pitchers is essentially the same.

offensive players, equating margins (for shot value). If instead basketball players were required to all shoot in a pre-determined sequential order, there would be little reason to expect margins to be equated.

The limited ability to transfer defensive resources across hitters should not be confused with ability to make optimal defensive adjustments for each hitter that do *not* involve transfers across players. Pitchers choose their type of pitches (i.e., fastball, change-up, slider) and the location of their pitches (i.e., inside, outside, high, low) based on each hitter's history at hitting such pitches. And fielders regularly shift in the field, based on the hitter's proclivity for hitting to certain spots, the game situation, and the type of pitches likely to be thrown. These common adjustments, however, do not involve moving resources from one hitter to another, but rather are made optimally for each hitter. Nor do they consume a limited defensive resource, thereby restricting the ability to make such adjustments for subsequent hitters. Consequently, there is no reason to expect such adjustments to equate margins across players: that is, there is no reason that hitter A and hitter B should perform the same after taking into account the optimal adjustments for each of them.

As mentioned earlier, the most significant adjustment a defensive team can make that affects subsequent hitters is to "pitch around" a good hitter. In particular, a team can avoid a hitter either directly, by intentionally walking him (putting him on first base without giving him the opportunity to hit), or indirectly, by attempting to pitch marginally hittable pitches on the border of or outside the strike zone (and consequently, increasing the chance that he will get a walk). And insofar as a walk is significantly more costly to the defense when first base is already occupied (i.e., if the previous player has walked), walking one hitter makes pitching around the next hitter even more costly, thereby providing a link across hitters.

Despite this link, such a strategy would not equate margins across different hitters for the statistics we consider, for two reasons. First, all of our statistics save for the walk and on base statistics are calculated exclusive of walks. For example, we consider the probability of hitting a home run in a given *at bat* conditional on the recent history of home runs per *at bat*, where *at bats* are defined by plate appearances by the batter exclusive of walks.¹³ So when a batter receives a walk, this will not count as an observation with our other three statistics, and as such, walks will not directly affect our measure of hotness. And second, pitching around a hitter is limited by the fact that in an average game situation, a walk is a well above average outcome for even the best of hitters (and correspondingly, a below average outcome for the pitcher). Consequently, in most settings, it is not optimal to pitch around a batter even if a weaker batter follows.¹⁴

While margins across players are therefore unlikely to be equated due to walks, better hitters

¹³Walks (as well as a few rarer outcomes) are excluded from at bats in the definition of these standard baseball statistics; this does not require any adjustment on our part from standard practice.

¹⁴And for the same reason, intentional walks are primarily employed only for specific, and fairly uncommon, settings – in particular, when a strong hitter is at bat, with men on base, with first base open, towards the end of a close game, and only when a weaker hitter is due to follow. This is precisely the setting when a base hit will do the most damage relative to a walk.

should be, and in fact are, walked more frequently. This allows us to use the prevalence of walks to test reactions to a hot hand: if better hitters are walked more frequently, we would also expect to see hot hitters, as measured by any of our four other statistics, walked more frequently as well – and in proportion to the inferred increase in their short-run ability in each statistic. Hence, we can test whether teams react to hot hitters with walks the same way they react to better hitters.

A team can also change pitchers, bringing in a better pitcher, a more rested pitcher, or a better matched-up pitcher, to face a better hitter. However, such opportunities are also limited, and we control for the quality of the pitcher that the hitter faces. The scope for such replacements are limited due to the rule that replaced players cannot re-enter a game. Given limited numbers of pitchers on a team, and the need to rest pitchers across games appropriately (due to the strain of pitching on one's arm), teams are very careful in their choice to replace pitchers. Typically, a team uses 2-4 pitchers in a game, with the starting pitcher expected to pitch most of the game, and relief pitchers used in the later innings if and when the starter tires or pitches poorly. Most pitcher replacements come at predictable times (based on the number of pitches the pitcher has thrown, how the pitcher has performed, the inning in the game, or the use of a pinch hitter) rather than based on the particular hitter to be faced. One important exception is at critical moments late in a close game, teams frequently do change pitchers in consideration of the particular hitter to be faced (and similarly, teams change hitters in consideration of the pitcher to be faced). The primary consideration for such changes is typically the platoon advantage – it is generally advantageous (disadvantageous) for the pitcher to be facing a hitter of the same (opposite) handedness as the batter. We control for the platoon effect (that is, whether the pitcher and batter are of the same or opposite hands) in all of our tests.

To summarize, in basketball, since offensive teams get to choose who shoots on each possession, optimal defensive strategy should involve equating shooting values across offensive players (i.e., a minmax strategy). In contrast, in baseball, all batters face pitchers, and one at a time, in a sequential order. This should not give rise to equating of values across hitters. Consequently, in contrast with basketball shooters, better hitters and pitchers should realize higher success rates than weaker hitters and pitchers. This is precisely what is observed in Figures 1 and 2. Moreover, the ability to pitch around better hitters on some occasions will afford us the opportunity to test whether and how much teams react to hot hitters, and in particular by compare this reaction to how they react to better hitters we can test whether teams appear to be making correct inferences on the magnitude of the hot hand.

2.2 Related Literature

Following GVT, there is a vast literature on streakiness in sports – primarily conducted by statisticians and sports researchers. Much of this research attempts to identify the hot-hand effect across

a variety of sports.¹⁵ Bar-Eli et al. (2006) provides a comprehensive review of prior literature. To summarize, they conclude that “*the empirical evidence for the existence of the hot hand is considerably limited*” and that while a number of the studies they survey claim to find evidence in favor of the hot hand, “*when one closely examines the results, demonstrations of hot hands per se are rare and often weak, due to various reasons...*”. Our results lie in stark contrast to these conclusions.¹⁶

Concerns related to endogenous responses in basketball have been noted from the very beginning. Yet, they have often operated in the periphery, whereas we argue they are fundamental to any empirical strategy. As mentioned earlier, GVT and others following them, attempt to address these concerns by obtaining similar results using free-throw performance and controlled shooting experiments. However, their attempts to do so lack power and are therefore unable to reject most plausible models of streakiness.¹⁷ More recently, both Arkes (2010) and Miller and Sanjurjo (2014) find strong evidence for a hot hand in controlled shooting environments that have significantly more power than prior tests, due to more data and the use of pooled data.

Our argument – that in equilibrium, one should not expect to see a hot-hand effect in the basketball shooting data even if it exists – is most closely related to Aharoni and Sarig (2012). They provide several pieces of indirect evidence that are consistent with both endogenous responses and the hot hand. Namely, that players with strong recent shooting performance (two or three consecutive successes) stay in the game longer and are more likely to take the next shot.¹⁸ Further, the presence of a player with strong recent shooting performance has a small positive influence on the shooting percentage of their teammates. Moreover, while teammates of a hot shooter exhibit higher shooting percentages (presumably due to the increased defensive attention on the hot shooter), the hot shooter’s shooting percentage, conditional on taking the next shot, remains unchanged. If instead the hot hand were a cognitive illusion then the shooting percentage of the hot shooter should decrease following the additional defensive attention.¹⁹

We start with a similar motivation – i.e., that defensive adjustments in basketball will equate shooting margins, thereby concealing shooting streakiness, but instead look for direct evidence of streakiness in another sport where we argue that such adjustments are less of a concern. As such, we look for and find, direct evidence of streakiness in outcomes in baseball, whereas Aharoni and

¹⁵See for example Larkey et al. (1989), Forthofer (1991), Adams (1992, 1995), Wardrop (1995, 1999), Gilden and Wilson (1995), Vergin (2000), Klaassen and Magnus (2001), Koehler and Conley (2003), Frame et al. (2003).

¹⁶In addition to this academic research, there is a nonacademic literature, focused on baseball research, conducted by a community of analytic sports fans. This literature reaches conclusions much in line with the academic literature. We discuss this literature in subsection 4.2.3.

¹⁷For example, Wardrop (1999) uses simulations to show that, for reasonable models of streakiness, the tests in GVT detect streakiness only 12% of the time.

¹⁸Bocskocsky et al. (2014) provide additional evidence that both offenses and defenses respond to recent shooting performance.

¹⁹Additional playing time and a higher propensity to shoot the next shot could also be explained by an incorrect perception (and hence suboptimal decision-making) on the part of players and coaches. For example, if a coach incorrectly believes that player A on the opposing team is hot and (suboptimally) devotes more defensive resources to that player, then less defensive resources will be devoted to the remaining players causing their shooting percentage to rise even if player A’s streak was purely random.

Sarig (2012) rely on indirect evidence in basketball. We then relate the magnitude of the streakiness effect to the degree to which defenses adjust to streaky behavior and test whether it is consistent with a hot hand fallacy. To our knowledge, we are the first paper to develop a test and present evidence on this question.

Other studies have attempted to address endogeneity by looking at sports in which endogenous responses are less of a concern. Supportive evidence in favor of the hot hand has been found in horseshoe pitching (Smith, 2003) and bowling (Dorsey-Palmateer and Smith, 2004), where the scope for endogenous responses is virtually non-existent. A common critique of these studies is that they are also generally underpowered and deliver (at best) weak statistical evidence for a hot-hand effect. Moreover, they have little to say about the magnitude of the effect and strategic implications. We differ from these papers in that we consider the hot hand and endogenous responses to it from an economic (i.e., equilibrium) perspective, providing a general theory of where one should and should not expect to find evidence for a hot hand, and we are able to relate the magnitude of our results to observed strategic responses and common perceptions.

Within baseball, statistical analysis has largely failed to identify support for streakiness (Swioff (1988), Gould (1989), Vergin (2000), and Albert and Bennett (2003)). Most of this analysis is conducted at the player or player-year level on hitting and involves conducting a runs test (or Wald-Wolfowitz test). Albright (1993) conducts regressions for the category of hits at the player-year level. He concludes that actual performance appears to be generated in a manner consistent with an i.i.d. process. Stern and Morris (1993) point out that Albright's analysis is biased against finding a streakiness effect and lacks power. The simulations we conduct in Section 6 are in agreement with Stern and Morris (1993).

We estimate regressions across the full panel of data and look at both batters and pitchers in a variety of statistical categories. By doing so, we obtain far more power than in prior tests, and in fact find strong evidence that the hot-hand effect is significant and strategically important in all ten of our statistical categories. Tango et al. (2007) also conducts tests on using panel data and concludes there is little to no hot-hand effect for batters. However, their tests are biased against finding a hot hand due to how they form expectations about a batter's ability. By appropriately accounting for this bias, the difference between their findings and ours is reconciled.

3 Tests for Streakiness

3.1 Data

We use data from Major League Baseball for the twelve seasons from 2000-2011, which is publicly available from Retrosheet.org. Each observation in the Retrosheet data set is an *event*, defined to be any occurrence that changes the state of the game, not including the pitch count. Most of the events correspond to a *plate appearance* (PA), which occurs when a batter completes his turn with a hit, walk, out or reaching base on an error. Each observation also includes roughly 90

variables that describe the state of the game prior to and following the event (e.g., batter, pitcher, score, outs, inning, base-runners, stadium, date, etc.). The data from 2000-2011 contains approximately 2 million observations.

3.2 Methodology

The term “hot hand” has been used loosely in the literature, typically without a precise definition. Here we discuss several different notions that have been associated with this term and with measuring streakiness.

3.2.1 What is a Hot Hand and Streakiness?

While there does not appear to be any consensus on how to define a hot hand, the literature has generally adopted a notion in line with short-term predictability. For example, does success/failure on a player’s last three field goals predict success/failure on the next field goal attempt (after controlling for all other predictive factors)?

One factor that should be controlled for is the underlying ability of the player in question. But it is also well acknowledged that players’ abilities change over time. Should such changes be considered a hot hand? Generally the literature has taken the view that low-frequency changes in ability are not to be associated with a hot hand. That Chris Davis hit 5 home runs in 2011 (playing only part time, presumably because his overall performance did not justify starting regularly), 33 home runs in 2012, and a league leading 53 home runs in 2013, is not taken as evidence of streakiness, but rather, is interpreted as evidence that he was a much better hitter in 2013 than 2011. In contrast, if a player were to experience, say, weekly, daily or hourly shocks to his ability, this would clearly fit the notion most seem to have for a hot hand. It is not clear, however, exactly where the line is to be drawn. While the hot-hand literature is vast, we are not aware of any clear delineation between what frequency changes in ability will be considered a hot hand, and what frequencies will be interpreted as a long-run change in ability.

We posit that when most players, coaches, and fans talk of a hot hand, they indeed are referring to predictability over a fairly high frequency – on the order of anything from within a game to 1-2 months. One frequently hear the claim that a player is hot this month (or week, or game, etc...), but one rarely hears the term applied to an entire season. A player performing much better for an entire season would instead typically be called a “much improved” player. Moreover, we posit that when players, coaches, and fans refer to a hot or cold player, and to strategic adjustments that accordingly should be made, they are referring to predictability given such recent performance. This is how we will refer to the hot hand as well; our general tests will ask whether recent history, after controlling for ability and other relevant factors, can predict outcomes.²⁰

²⁰In our standard tests, we will use both prior outcomes, and future outcomes, outside of a window “near” a given event, to control for ability. We also run the same tests using only prior outcomes to control for ability, so that our regressions are indeed predictive. Both sets of tests yield similar results, as we discuss later.

In our tests, we will typically take the length of the recent history to be 25 at-bats or plate appearances. For the typical hitter, this will be on the order of 5 or 6 days. This is a longer history than much (but not all) of the literature. We discuss this choice in more detail in Sections 3.4 and 6. In that discussion we point out that restricting the history to be very short (i.e., last 3 shots as in GVT) significantly and arbitrarily restricts the power of tests – even in a world of strong hot hands, it is very hard to know whether a 3-shot run was due to luck or a hot hand.

We will also depart from much (but not all) of the literature, in using panel data. In the literature, most tests are performed at the player level.²¹ The advantage of performing tests at the player level is that they do not require an additional distinction between the player’s ability and whether the player is hot or cold. Under the null hypothesis, the ability of the player on which the test is performed is assumed to be constant throughout the sample period (often a year), and the test is whether recent past performance predicts the outcome of future attempts. As pointed out by others and discussed in Section 6.2, due to the small sample size, this approach generally lacks power to reject simple, yet plausible, models of streakiness.

Using panel data affords a much greater number of observations, which gives us the power necessary to identify the presence of streakiness. However, it also requires us to distinguish between a *good* player and a *hot* player. That is, we must control for each player’s ability when trying to detect any deviations from it.

The association of a hot hand with predictability introduces an issue in interpretation, that is also present but generally unacknowledged in other papers in the area. In particular, predictability may derive from short-term changes in ability, or from learning about longer-term ability. We refer to the former as “streakiness” and the latter as “learning”. We use the term “hot hand” synonymously with short-term predictability, which encompasses both streakiness and learning.²² After finding strong evidence for predictability, we will then proceed in Section 4.2.1 to decompose this effect into the two components of streakiness and learning. Our results vary across the different statistics, but in general we find that streakiness accounts about half of our hot-hand results, with learning attributable to the other half of the effect.

3.3 Regression Model

We estimate a (reduced-form) regression model (both OLS and logistic) for ten different *outcome statistics* of particular interest (five for batters and five for pitchers). Namely, whether the plate appearance resulted in a hit, homerun, strikeout, walk as well as whether the batter reached base. When considering each of the first three outcomes (hit, homerun and strikeout), we drop all plate

²¹Examples of recent studies employing panel data include Arkes (2010), Aharoni and Sarig (2012), Arkes (2014) and Boscokosky et al. (2014).

²²One might object that behavioralists often have only short-term streakiness in mind when discussing the hot hand. But at the same time, most prior tests for the hot hand have been tests for predictability, and common usage of the term (by players, coaches, and fans) seem to encompass all effects of predictability – i.e., to refer to what can be inferred about upcoming outcomes from recent play.

appearances that did not result in an official *at bat*.²³ At bats are plate appearances excluding walks as well as several other rare occurrences in which the plate appearance terminates without the batter attempting to earn a hit (e.g., when the batter is hit by a pitch or hits a sacrifice bunt). Letting Y_{it} denote an indicator for whether the outcome Y occurred in player i 's attempt at time t , we estimate the following OLS and logistic regression equations:

$$Y_{it} = \alpha + \gamma \text{State}_{it} + \beta X_{it} + \epsilon_{it} \quad (1)$$

$$\log\left(\frac{P_{it}}{1 - P_{it}}\right) = \alpha + \gamma \text{State}_{it} + \beta X_{it} \quad (2)$$

where $P_{it} = \Pr(Y_{it} = 1)$, State_{it} is a measure of how well player i has performed recently and X_{it} denotes a vector of control variables, which includes the long-term ability of player i , the ability of the pitcher or batter he is facing, and other explanatory variables. For each outcome, we estimate the model in which i refers to a batter in a given year and the model in which i refers to a pitcher in a given year.

The coefficient of interest, γ , indicates whether short-run performance predicts the outcome. If it does not, then we would expect $\gamma = 0$. While the standard test for a hot hand in the literature is formulated around whether gamma is positive or not, a positive coefficient could indicate streakiness or learning or both. In section 4.2.1 we will give evidence for both: our results indicate that predictability arises from both channels. In contrast, a negative coefficient would be indicative of reversals in outcome.

3.4 Measuring State and Ability

In our baseline regression model, we estimate the state of a player using his recent success rate in his last L attempts.

$$\text{State}_{it} = R_{it}^L \equiv \frac{1}{L} \sum_{k=1}^L Y_{i,t-k}. \quad (3)$$

In selecting the history length L , one encounters the following tradeoff. If the state does not change over the period, a longer period will provide a more accurate identification of the state. A hitter who has three hits in his last three at-bats might be hot, or he might have been lucky. More at-bats in the history can help distinguish between the two. However, if the state changes significantly over the length of the history, one will be using a history that is less relevant for measuring the current state and more relevant for measuring long-term ability. For the majority of our analysis we use $L = 25$ (and drop the L superscript). This history length corresponds roughly to performance in the 5 most recent games for a batter and takes roughly a week for a “regular” batter. To account for this, we also require that these 25 most recent attempts occurred within

²³This is consistent with the standard baseball statistic of batting average.

the last 10 days so as to ensure the measure captures “recent” performance. We arrived at $L = 25$ because this history length seems in line with conventional notions of how long a typical streak lasts (i.e., on the order of a week or two). We also report estimates using the full data set for two alternative lengths $L=10$ and $L=40$ for which we obtained qualitatively similar results. We consider several additional methods for measuring a player’s state, reported in Section 3.6.1, for which we also obtain similar results. Note that in contrast with our use of $L = 25$, much of the hot-hand literature, including GVT considers histories of $L = 1, 2$ or 3 and only those that occur within a game; i.e., is a shooter more likely to make a shot after making/missing his last shot in that particular game. But in a setting where a “hot hand” last much longer than 1-3 attempts, such a test is poorly designed, with particularly low power. For most reasonable models of streakiness, success in the last outcome, or even the last 3 outcomes, will be sufficiently noisy that a longer history would serve better in identifying potential hot states (see Wardrop (1999) and Section 6 below for more on this point).

We consider a variety of methods to control for the player’s *ability*. In the baseline model, we do so by calculating the player’s rate of success during the season not including a “window” of W attempts before and after the current attempt.

$$\text{Ability}_{it} = \frac{1}{N_i - 2W} \times \sum_{s \notin [t-W, t+W]} Y_{is} \quad (4)$$

It is important not to include the dependent outcome variable in the measurement of ability. Moreover, we want to exclude nearby outcomes as well, since under the alternative hypothesis of a hot hand, these outcomes will be positively correlated with the current outcome. Failure to exclude the current outcome (the dependent variable) from the measurement of long-term ability will lead to a downward bias in the estimate of the hot-hand effect under both the null and alternative hypothesis. Failure to exclude nearby outcomes will lead to a downward bias under the alternative hypothesis.²⁴

To see the latter effect, consider the alternative hypothesis of a hot hand. Under this hypothesis, the outcome of the current at bat (insofar as it is indicative of the current state) will predict near-future outcomes. Including these near future outcomes in ability will consequently lead to an overestimate on the effect of long-term ability, and an underestimate of the hot-hand. To mitigate these potential biases, under both the null and the alternative hypotheses, we measure ability using overall season performance excluding a window – typically of 50 at-bats before and 50 at-bats after – around the current outcome. We chose $W = 50$ in an effort to reduce any lasting effects of the player’s (potential) streakiness (recall that we use $L = 25$ for the baseline model), while at the same time balancing the need to maintain a statistically meaningful estimate of the players ability based

²⁴We emphasize this point because some recent related studies do not exclude the outcome from the ability measurement (see e.g., Arkes (2010); Chen et al. (2014); Bocskosky et al. (2014)).

on $N_i - 2W$ observations.²⁵ In this vein, we drop observations in which $N_i - 2W < 100$, though our results are strengthened if these observations are included. We also consider several other ways to control for ability (see Section 3.6.2), for which we obtain similar results. For example (and in order for our results to be predictive), we consider an ability measure using only prior outcomes.

3.5 Control Variables

In addition to controlling for the state and ability of the player, additional control variables are included in the regressions. In particular, the ability of the opposing batter/pitcher, the stadium, time trends, and whether the pitcher and batter were of the same hand. Ability of the opposing batter or pitcher (player j) was measured by computing the success rate of all batters (pitchers) who faced pitcher (batter) j in that year excluding the current at bat. For example, we use the batting average of batters who faced pitcher j over the course of the year excluding the current at bat. We control for the stadium and time trends by including stadium \times year fixed effects. We also include a dummy variable to control for whether the batter and pitcher were of the same hand. The dummy variable is equal to 1 if the pitcher's throwing arm is the same as the side from which the batter hit.

As we have already argued, the most significant potential endogenous defensive responses can be controlled for in the data. For example, a team may choose to "pitch around" or intentionally walk a good hitter, or it may bring in a better pitcher to face a good hitter. In both cases, these responses are accounted for in our analysis. First, we drop all walks when predicting hits, homeruns and strikeouts. Second, in the regression analysis, we control for both pitcher ability and the platoon effect. In Section 5, we explore the magnitude of this response in greater detail.

Our data set provides a wide range of additional situational variables that one might expect to predict the success of a particular outcome (e.g., inning of at bat, number of outs, number of runners on base or in scoring position). However, similar to Albright (1993), these variables were not found to be statistically significant nor do they alter the predictions of the model and are generally omitted from our reported results.

3.6 Robustness

In addition to the baseline regression, we consider a number of alternative specifications. Most of these involve using other measures for state and ability.

3.6.1 Alternative Measures of State

We consider several alternative measures of state. The first is to use dummy variables to indicate whether a player, based on recent performance, appears to be hot, cold, or neutral, *relative to his*

²⁵Using this measure will result in unbiased estimates if the true data generating process is i.i.d. Further, it is consistent asymptotically if as $N_i \rightarrow \infty$, we let $W \rightarrow \infty$ at a rate slower than N_i so that $N_i - 2W \rightarrow \infty$.

own normal performance. To do so, we use an estimate in the form of $\widehat{State}_{it} = [Hot_{it} \ Cold_{it}]$. We consider both additive and multiplicative cutoffs here:

- *Additive Cutoffs:* A player is considered hot (cold) if his recent success rate is A (B) percentage points above (below) his long term success rate or *ability*. That is,

$$Hot_{it} = 1 \{R_{it} \geq Ability_{it} + A\}$$

$$Cold_{it} = 1 \{R_{it} \leq Ability_{it} - B\}$$

where A , B are chosen so that the average player is hot, in 5% of all at bats and cold in 5% of all attempts. For this, and each of the specifications using dummy variables below, we obtained similar results for alternative using 2.5% and 10% thresholds. We reported these in the categories of batting average and on-base percentage for batters in Tables B.3 and B.2.²⁶

- *Proportional Cutoffs:* A player is considered hot (cold) if his recent success rate is $q > 1$ ($r < 1$) times above (below) his ability. That is,

$$Hot_{it} = 1 \{R_{it} \geq q \cdot Ability_{it}\}$$

$$Cold_{it} = 1 \{R_{it} \leq r \cdot Ability_{it}\}$$

where q , r are chosen so that the average player is hot in 5% of all at bats and cold in 5% of all at bats.

As a convention, we maintain the same criterion across both batters and pitchers, despite the fact that outcomes which are considered “good” for batters, such as hits and homeruns, are considered “bad” for pitchers and vice versa (strikeouts). The advantage of this convention is to maintain the same sign prediction for being hot (or cold) across all outcomes. That is, regardless of whether an outcome is good or bad for the player, if the outcome has occurred more frequently recent past, than if players are streaky, we would expect the outcome is more likely to occur in the current attempt and hence obtain a positive (negative) coefficient on the hot (cold) dummy variable. In the category of home runs, R_{it} takes a value of zero in more than 40% of attempts. It therefore seems unnatural to classify a batter as cold in this category, and we omit the dummy variable “cold” from the homerun regressions.

The second alternative strategy we employ is to compute how the player’s current recent history, R_{it} , compares to his overall distribution of histories. In this case, we construct an estimate of the player’s state, which corresponds to the fraction of the time that his recent history is above the distribution of his histories in the opposite half of the season. Letting $\mathcal{S}(t) \in \{1, 2\}$ denote whether

²⁶Results for other categories are available upon request.

the attempt t occurred in the first or second half of the season, we let

$$\widehat{\text{State}}_{it} = \frac{1}{|\{s \notin \mathcal{S}(t)\}|} \sum_{s \notin \mathcal{S}(t)} 1\{R_{it}^L > R_{is}^L\} = \text{freq}\{R_{it}^L > R_{is}^L, \forall s \notin \mathcal{S}(t)\} \quad (5)$$

For example, a value of 0.7 means that the player's recent history in attempt t is better than 70% of all of histories for that player's attempts in the opposite half of the season. The higher is the value, the better the recent performance has been relative to the player's performance in the opposite half of the season. Making use of this statistic for the player's state, we first run a regression of the form given by equation (1). We then construct dummy estimates of the form given above using the following method.

- *Distribution Based Cutoffs:* A player is considered to be hot in attempt t if R_{it} is in the right tail (top 5%) of the distribution from the opposite half of his season and cold if R_{it} is in the left tail (bottom 5%).²⁷

3.6.2 Alternate Measures of Ability

An alternative way to control for player ability is by using player fixed effects. We estimate the following regression model:

$$Y_{it} = \alpha_i + \gamma \text{State}_{it} + \beta X_{it} + \epsilon_{it} \quad (6)$$

Estimating (6) using standard techniques generates a downward bias on γ (under both the null and alternative hypotheses) due to the lagged dependent variables in the State variable.²⁸ The intuition for the bias is essentially the same as the logic given above for why a window is needed.²⁹ Therefore, we have also considered several additional measures to control for the players ability including: the player's success rate in the previous year, the player's success rate in each of the previous three years, the player's success rate over the entire sample omitting the season in which the observation takes place, and the players' success rate over the entire sample including the season in which the observation takes place. While the latter suffers from the overlap bias, the estimator is still consistent and the number of observation is significantly larger than at the season level, which mitigates the bias. For each of these measures, we obtain regression estimates similar to those when using the windowed measure of ability.³⁰

²⁷It is necessary to use the opposite half of the season for this criteria in order to avoid the overlap bias. To see this, let t_i denote the time at which i 's recent success rate is at its maximum point and assume for simplicity that t_i is unique. Since $R_{it+1} < R_{it}$, it must be that $Y_{it} = 0$. In other words, if a player has been very successful recently relative to the rest of the *entire* season, then in his next attempt the unconditional probability of success must necessarily be lower.

²⁸See, for example, Wooldridge (2010, p 373).

²⁹Including a player (player-year) fixed effect is identical to using the lifetime (annual) batting average of the player as an ability control.

³⁰Results available upon request.

4 Evidence of Streakiness

4.1 Results

Regression estimates are located in the Appendix. Standard errors are clustered at the player level. Tables A.5-A.14 contain the estimates for each of the our 10 statistics, for the main specifications discussed above. The results are quite striking. For all 10 statistics that we consider, the estimate for γ , the coefficient for players' recent success rate, R_{it}^L , is statistically significant at a 0.1% confidence level for both our OLS and logistic specifications (Columns (1) and (2) in each table). Thus, we find strong evidence for the hot hand across all 10 statistics. We also find similar results across our alternative measures of hotness. The coefficient on all recent history variables are also statistically significant at a 0.1% level for all 10 statistics and for all three specifications given in columns (3)-(5), where we consider a measure of streakiness based on each player's own distribution of outcomes (column (3)), a multiplicative cut-off for determining who is hot and cold (column (4)), and an additive cut-off for determining who is hot and cold (column (5)). The only specification where some of our results are not significant at a 0.1% level is when we use distributional cut-offs to determine who is hot and cold (column (6)). Even in that specification, we find significance for our hot and cold variables at a 5% confidence level in all 10 statistics and at a 0.1% level in 8 out of the 10 statistics.

Not only are our results strongly significant statistically, they are also significant for strategic considerations. For example, as we discuss in Section 7.2, the magnitudes we observe are the right order of magnitude to rationalize common managerial decisions made with regard to streaky players.

For all these tables, column (1) gives are base OLS regression, with the coefficient on state representing the increase (in percentage points) in the likelihood of the outcome given a one percentage point increase in R_{it}^L . For example, column (1) in Table A.7 indicates that a .100 difference in a batter's recent on-base percentage translates to a .00628 increase in the probability of getting on base in the next plate appearance. Thus the difference between a recent on-base percentage of .100 and .500 (representing a recent cold run and hot run respectively) translates to an increased probability of getting on base of $0.4 \times 0.0628 \approx 2.5\%$, or 25 OBP points. This magnitude is quite significant from a strategic perspective: Table A.2 indicates a cross-sectional standard deviation of .0299 in on-base percentage across hitters throughout the sample; thus 25 OBP points corresponds to approximately $\frac{5}{6}$ of a standard deviation in hitter ability. Put differently, the difference between a hot player and the same player when cold corresponds to approximately the difference between a 50th percentile hitter and an 80th percentile hitter.

The coefficient on the logistic regressions in column (2) are less easily interpreted. However, we verified that they imply marginal effects consistent with those in the linear regressions. Similarly, columns (4)-(5) indicate that the difference between a hot and cold hitter as defined in these two specifications (where 5th percentile hot and cold cutoffs are defined using additive and propor-

tional measures respectively) is about 28 OBP points for both specifications. Columns (3) and (6) present the results using the distribution-based measure given in equation 5. Here, we find a somewhat smaller estimate for the difference between a hot and cold hitter.³¹ For example, using the distribution-based cutoff in on-base percentage for batters (i.e., column (6) of Table A.7) yields approximately a 16 OBP difference between hot and cold hitters.³²

The magnitude of the effects are quite similar for pitchers. Looking at OBP against pitchers in Table A.8, column (1) indicates that an .100 difference in a pitcher’s recent on-base percentage allowed translates to a .00696 increase in the probability of the next hitter to face the pitcher getting on base. Thus the difference between a recent on-base percentage allowed of .100 and .500 translates to an increased probability of allowing the next hitter to get on base of $0.4 \times 0.0696 \approx 2.8\%$, or 28 OBP points. Table A.2 indicates a cross-sectional standard deviation of .0284 in on-base percentage allowed across pitchers in the sample; thus 28 OBP points corresponds to approximately one standard deviation in pitcher ability, or approximately the difference between a 50th percentile pitcher and an 84th percentile pitcher. As with the hitters, Columns (4)-(5) find a similar difference between hot and cold pitchers using 5th percentile additive and proportional cutoffs, finding a difference of approximately 30 and 26 OBP respectively. Also similar to hitters, Column (6) indicates that using the distributional cutoff for hot and cold pitchers yields somewhat smaller estimate, producing approximately a 20 OBP difference between hot and cold pitchers.

As another example, Column (1) in Table A.9 indicates that if a player hits home runs in an additional 10% of his last 25 at bats – i.e., hits an extra 2.5 home runs in the last 25 at bats, the likelihood of hitting a home run in the next at bat increases by .75%. This compares with the average batter who hits a home run in 3.0% of his at bats, with a standard deviation of 1.7% across players. Similarly, players classified as hot by the additive cutoffs in column (5) similarly have a .79% increase in their likelihood of hitting a home run in their next at bat relative to when they are in a normal state.³³ The effect with pitchers and home runs, in contrast, is somewhat weaker. For example, Column (1) in Table A.10 indicates that if a pitcher gives up home runs in an additional 10% of his last 25 batters faced, the likelihood of giving up a home run to the next batter increases by .33%. This smaller streakiness effect for pitchers, however, can be understood in the context of a significantly less cross-sectional variation across pitchers; the standard deviation in home run rate allowed across pitchers is .098.

One useful way of organizing and summarizing our results across all statistics is to compare the difference in performance between a hot and a cold player to that of a one standard deviation in long-run average across players for the statistic. Using our 5% additive cutoffs in column (4), we

³¹One can multiply the coefficient in column (3) by 0.9 in order to obtain an estimate for the difference between a hitter in the 95th percentile of their own distribution from the opposite half of the season to one in the 5th percentile.

³²A likely explanation for the smaller estimate under this measure is due to the fact that the distribution-based measure is noisier in that only the most recent 25 attempts and opposite half of the season are used to determine the players state as opposed to all attempts outside the window (as in the other measures).

³³It is difficult to classify a batter as “cold” in this category given that in more than 40% of all at bats a batter has no home runs in his last 25 at bats.

report these estimates in Table A.15. The difference between a hot and cold player inferred from recent history is ranges from roughly one-half to one standard deviation in the variation across overall player ability. Averaged across all 5 statistics, we find the difference between a hot and cold player corresponds to 0.84 and 0.68 of a standard deviation in long-run average for batters and pitchers respectively.

4.1.1 Additional Robustness Tests

Additional tables in Appendix B contain a number of robustness checks. Specifically, Table B.1 contains estimates for the regression equation (6), which includes batter fixed effects as an alternative control for ability. In all the five categories, the coefficient on the state is positive, and it is statistically significant in four of the categories.

Tables B.2-B.3 use 2.5% and 10% thresholds (as opposed to 5%) in order to classify a batter as hot or cold (see Section 3.6.1). For brevity, results are reported for batters in two categories (hits and on base). The statistical significance of the estimates is similar to the 5% case. When using the 2.5% (10%) threshold, the magnitude of the coefficients are moderately larger (smaller) than the 5% thresholds. One interpretation is that the 10% threshold leads to a higher frequency of Type I errors (incorrectly identifying the player as hot or cold), thereby attenuating the estimated size of the effect.

Finally, Tables B.4-B.7 contain estimates when using a longer ($L = 40$) and shorter ($L = 10$) history lengths in the estimate of the players's state as given by equations (3). The sign of the estimates and the order of magnitude is the same for all 10 statistics both for the longer and shorter length history. The longer (shorter) history length results in somewhat stronger (weaker) estimates (presumably due to the learning effect discussed in Section 4.2.1).

4.2 Interpretation

4.2.1 Streakiness and Learning

We find statistically and strategically significant results for our “hot-hand” variable across all ten statistics we consider and all specifications. These results indicate that even after controlling for player ability using long-term performance, short-run performance has significant predictive power on the outcome. This contrasts with prior hot-hand tests, which do not find predictability in short-run performance.

It is important to realize that part of our findings should be interpreted as learning rather than streakiness. In particular, our measurement of long-term player ability, while unbiased, is a noisy measurement given finite observations. Consequently, even without streakiness, there is further information to be learned about long-run ability from the short-run history, which would lead to a positive coefficient on the short-run performance, even absent any streakiness. Note that much of the literature does not make this distinction – i.e., most other hot-hand studies look for any effect

of the recent short-term history after controlling for some measure of long-run ability, and purport to find no effect.³⁴

In this subsection we decompose our “hot hand” results into a learning effect and a streakiness effect. In considering such a decomposition, it is important to note that both components are typically what is considered in hot hand tests, and moreover it is the combined effect of the two that should govern strategic decisions. For example, a manager in deciding who to play in the next game should optimally be taking into account both learning and streakiness inferred from a player’s recent performance in assessing his likely performance for the next game.

There are several manners to disentangle learning from streakiness. First, under the null of no streakiness, the magnitude of this coefficient relative to that on our long-run variable should in expectation equal the ratio of the number of observations for the short-run variable relative to the number of observations for the long-run variable. Hence, we can compare our coefficients to the predicted value given learning under the null, and attribute any difference to streakiness.

Second, we can change our definition of the long-run ability variable to also include the most recent 50 prior outcomes, thereby overlapping this variable with the range for short-term performance. Then there will be no new information in the short-term performance measurement that is not already included in the long-run ability variable. Note that with this specification, the coefficient γ will understate the predictive ability of short-term performance insofar as the learning effect will be included in long-run ability.

And third, we can compare the coefficient we find on recent performance to the coefficient on a similar regression, where instead of using short-term performance to predict the next outcome, we use it to predict the 11th outcome, or the 26th outcome, etc. Consider our standard regression, but now for outcome t , instead of defining short-term performance to be outcomes from $t - 25$ to $t - 1$, instead let it be from $t - 25 - l$ to $t - 1 - l$, where l is some positive lag. The larger l , the greater the distance from the short-term variable and the outcome, and hence the smaller should be the effect of any streakiness. Learning, on the other hand, should not be affected by l , because provided that $l \leq 25$ (which ensures that the short-term performance variable does not overlap with the long-run ability variable), the short-run performance variable will have the same amount of new information (i.e., 25 additional observations not included in the control variable for ability). Consequently, we would expect γ to fall with l given streakiness, but also to remain positive, given the presence of learning.³⁵

All three of these tests yield similar results, demonstrating the importance of both streakiness

³⁴That prior studies find no effect on the short-run history in the face of such learning demonstrates the low power of such tests. That is, not only are they not finding streakiness, they are not finding any evidence of learning from the short-run history, which should exist even absent a hot hand. The likely problem here is that many prior studies use a short-run history that is too short; both too short to identify whether the player is hot, and too short for much learning.

³⁵We cannot consider $l > 25$ in the context of our basic specification, insofar as once l exceeds 25, then the short-term performance history will overlap with the long-run ability window, leading to a different interpretation of our results.

and learning in our findings. First, we simply compare the coefficient on the recent performance variable to what we would expect it to be under the null of no streakiness, given the presence of learning. Assuming a player's ability does not change over a sample period, and if all observations yield the same amount of information about ability, the weight that one would put on two discrete samples of outcomes in forming inferences about ability should be proportional to the number of observations in the two samples. Hence we can use our coefficient on long-run ability to project what the coefficient on the recent history would be under the null of no short-term variation in ability. By construction, the recent performance variable comes from 25 observations. The number of observations that goes into the average hitters' (pitchers') ability measure is given by 448 (593) for statistics that use plate appearances, and 402 (551) for statistics that use at-bats.

Using our base specification across our 10 statistical categories, Table A.19 gives the estimated coefficients for ability and recent performance, the expected coefficient on recent performance under the null hypothesis of learning absent streakiness, the ratio of the expected short-run coefficient under this null of learning to the observed coefficient, and the F-statistic from a Wald test as to whether the estimated γ can be explained by the null hypothesis of learning absent streakiness. In 9 of the 10 statistics, we reject strongly the hypothesis that the two coefficients are equal. For the batting average statistic for batters, the coefficient is larger than it would be under the null but the hypothesis can only be rejected at a 12 percent confidence level.

As this table indicates, the null of learning can explain from 25.6% of the observed short-run coefficient (for on-base percentage allowed by pitchers) to 73.9% of the observed short-run coefficient (for hitters batting average). On average, learning accounts for 54.0% of the coefficient averaged across our 5 hitting categories, and 38.6% of the coefficient averaged across the 5 pitching categories, or 46.3% of the coefficient averaged across all 10 statistics.

Our second test compares the coefficient on the recent history for our base case regressions for all ten variables with the same coefficient on regressions that instead includes the most recent 50 outcomes in the long-run ability variable (rather than excluding these as in the base case). Hence, the short-run performance is included in the long-run ability variable, and there should be no additional learning from the short-run performance.

When the recent history is included in the long-run ability variable (overlapping with the short-run history variable), the coefficient on the short-run history falls for all 10 statistics. This is as expected, as the short-run history variable previously included both learning and streakiness; but including these observations in the long-run ability should effectively remove the learning effect. The ratio of the coefficient on the short-term history when recent history is included in the long-run ability variable to the same coefficient when recent history is not included in the long-run ability variable (i.e., our base case) can be loosely interpreted as the fraction of our initial short-run history effect that is attributable to streakiness.

Results for this streakiness-learning decomposition in Table A.20 are similar to those in Table A.19. Averaged across all 10 statistics, we find that learning accounts for a little more than half

of the effect (56.1%) and consequently, streakiness for a little less than half of the effect. Streakiness is more prominent relative to learning with pitcher statistics and less prominent relative to learning with the pitchers statistics, and less so with the batter statistics.

For our third test we consider lagging the short-term performance variable by l . Thus a lag of $l = 25$ means that instead of using observations from $t - 25$ to $t - 1$ to predict outcome t , we use observations from $t - 50$ to $t - 26$ to predict outcome t (using all other controls used previously, including the same base case specification for long-term ability). Table A.21 reports the short term performance coefficient, for our base case specification, together with $l = 5$, $l = 10$, and $l = 25$, and the ratio of the coefficient for $l = 25$ with the base case ($l = 0$) for each of our 10 statistics.

By lagging the short-term history, the effect of streakiness should be diminished, whereas there should be no effect at all on learning. The greater the lag, the smaller the streakiness effect: With a lag of only 5 at bats, we still expect to observe a high degree of streakiness – a hitter who was hot over a 25 at-bat history is still likely to be hot 5 at-bats into the future. Even a lag of 25 at-bats should still display a significant degree of streakiness, though it should be smaller than with a lag of 5 at-bats. Since learning should not be affected by this lag, a decreasing coefficient on the short-term history as the lag is increased is a clear indication of streakiness independent from learning.

Table A.21 demonstrates results very much in line with the presence of both streakiness and learning. As the lag increases, the magnitude of the coefficient of the short-run history falls, as expected. The average coefficient given lags of 5, 10 and 25, as a fraction of the base case (lag of 0), is 88.7%, 83.4%, and 74.1% respectively. This decrease with the lag is indicative of streakiness; the predictive power of the short-run history falls as the current at-bat is further away. That it falls to only 74.1% with a lag of 25 is indicative both of learning (as the two prior tables demonstrate that around 50% of the total effect is accounted for by learning), and the persistence of streakiness over more than 25 attempts.

Taken together, Tables A.19, A.20, and A.21, tell a consistent story. Roughly half of the effect that we find on the short-run history for our average statistic is attributable to learning, and half is attributable to streakiness. Streakiness relative to learning accounts for a little more for the average pitching statistic and a little less for the average hitting statistic. Streakiness fades over time, but is still relatively important with a 25 at-bat lag. Finally, it is worth emphasizing again, that both the streakiness and the learning effects that we have separated in this section contribute to what common parlance would refer to as a “hot hand”. That is, both effects contribute to the predictability of recent performance, and both are strategically relevant. A player or coach trying to predict current outcomes based on recent history should be taking both components into account, and for such predictive purposes this decomposition might be largely irrelevant.

4.2.2 Predictability

All of our analysis uses data from future attempts to control for player ability. For example, in our base specification, we use the player's success percentage in the current year excluding the previous 50 attempts and the next 50 attempts, which will generally include the player's performance in future attempts (e.g., 51,52,...). From the perspective of sports strategists, coaches, and fans who are interested in predicting outcomes such measures of ability will not be available and one has to rely on a measure based only on past performance. Based on the discussion in the previous section, it should be no surprise that our results are strengthened (i.e., the coefficient on the state variable in our primary specification increases) if we exclude the use of future attempts in controlling for player ability. This, of course, occurs because there are fewer observations from in the control for ability relative to the number of observations in the short-term performance and hence more learning from short-term performance.

Conversely, our primary specification uses only the current season of data to control for player ability. A better predictor of a player's long-term ability would include an (optimally) weighted average of the player performance in the current season as well as previous seasons.³⁶ We have run a variety of robustness checks along these lines. In general, the more data points and degrees of freedom we allow for in controlling for player ability, the less there is to learn about the player from the most recent attempts and the smaller is the coefficient on the state variable. While the learning component shrinks, the streakiness component persists. For example, even when we control for ability using more sophisticated measures—say, an optimally weighted average of performance from: the current season (prior to the current attempt), the previous season, two seasons ago, and three seasons ago—we still find a positive and statistically significant coefficient on the state variable.³⁷

4.2.3 Sabermetric Studies

A number of studies in the sabermetric literature find little to no evidence of a hot hand in baseball.³⁸ However, we have found that these studies typically have a flawed measurement of ability, leading to a downward bias in the measurement of the hot hand. We will discuss one such study in detail here (Tango et al., 2007, Chap 2) because it is similar in conception to our work, but also representative of this drawback.

As in our study, Tango et al. (2007) tests for the hot hand by asking whether recent performance predicts future outcomes, controlling for player ability and other factors. In this study they compare how hot players and cold players perform relative to their “expected” performance. Performance

³⁶Since the player's long-term ability is likely to change over time, the optimal weights are higher for performance in more recent years.

³⁷Estimates from these regressions are available upon request.

³⁸The sabermetric community is a large informal community dedicated to empirical baseball analysis. (The name sabermetrics being derived from the acronym for the Society for the Advancement of Baseball Research.)

is measured with weighted on base average (wOBA) (an ability measure devised by one of the authors, Tom Tango), which differs somewhat from all of our measures, but should not behave that differently from our statistics. They find very slight predictive power; wOBA for the next game is .369 for hot players compared to a predicted value of .365, and is .330 for cold players, compared to a predicted ability of .337. Thus, they find some evidence for a hot hand, but of a smaller magnitude than we find for all of our statistics and tests.

We believe this difference, however, is attributable to two flaws in their construction of “expected” performance. First, they construct this measure using actual overall outcomes for the player in question in the prior, current, and following season, thereby including the dependent variable, as well as adjacent at-bats, in this measure. And second, their measurement of expected outcome is biased and high for above average players and biased and low for below average players, and moreover, the former (latter) are over-sampled in their hot (cold) set of players.

To understand this second drawback requires some detail. Their study define hot and cold players by a fixed five percent threshold in performance over the past five games. The same threshold is used for all players, hence good players are over-represented in the hot sample, and bad players in the cold sample. They then estimate player ability by looking at overall performance in the three season span including the prior, current, and following season. In this manner they calculate the average ability of players in both their hot and cold sample. They then compare how the players in the hot and cold sample actually perform relative to these expected abilities, adjusting for many of the same controls that we consider – quality of pitcher faced, ballpark, etc.

The problem here, however, is that this average ability measurement is not an unbiased predictor of how players in each sample should perform. For each player they simply use the average performance over the 3 year span as the basis for their expected performance; they are not employing a fitted value from a regression. This method might be close to unbiased for an *average* hitter. But the best predictor for a hitter who has performed better (worse) than league average over a given span of time will be less than (greater than) than his observed performance over this span. This regression to the mean follows naturally from any noisy learning environment, where good performance may be due to ability or may be due to luck, and is readily verified to be true in baseball data. Hence a sample that is over-represented with good (bad) players, as their hot (cold) sample is constructed, will over-estimate (under-estimate) expected performance. This in turn will lead to a downward bias in performance attributed to a hot hand.

To give a rough quantification of this bias, we regress player performance in a given year on a 3-year history of average past performance and a constant. Depending on the statistic, we find a weight on the order of about .7 on performance and .3 on the constant (that would reflect the league average).³⁹ Since the difference in actual wOBA between the hot and cold sample used in

³⁹While we do not have specific numbers for the authors chosen statistic of wOBA, we posit that results would be not differ much, given the similarity of wOBA to several of our statistics. Moreover, if they have chosen to include in their study players with relatively few at-bats in their 3-year window, the likely coefficient on the constant would be larger than .3, leading to a larger bias.

their study is 28 wOBA points, the overall size of the bias is on the order of $28 * (.3) = 8.4$ wOBA points for the hot and cold players combined. Compared to their finding of 11 wOBA points for the sum of the hot and cold players, this second bias alone would effectively double the magnitude of their results, bringing them much closer to the magnitudes that we find. If we adjust for the likely effect of their first bias as well, their study ends up finding a hot hand similar in magnitude to our findings.⁴⁰

5 Tests and Evidence for Endogenous Responses

The prior section documents strong evidence for a significant hot hand in baseball. The fact that better hitters are walked more frequently by defenses allows us to ask about team inferences and responses to such streakiness. If players are walked more due to a higher long-term or permanent ability, then they should also be walked more when they possess a short-term component to their ability (streakiness) that is inferred to make an equivalent contribution to the outcome. In this section we use walks to ask if defenses respond to streaky hitters. And if so, is their response consistent with them drawing a correct inference about the magnitude of the effect? Or, do they tend to over-react to recent performance in line with a hot-hand fallacy?

As discussed above, the most significant adjustments a defensive team in baseball can make across hitters is to “pitch around” good hitters; that is, to either partially avoid the hitter by throwing pitches outside the strike zone, thereby increasing the likelihood the batter is walked, or to completely avoid the hitter by intentionally walking him. It is well known that better hitters are pitched around and intentionally walked more frequently. This is especially true for power hitters (i.e., those that hit a lot of home runs). Thus, we would also expect that hot hitters (and especially those hot in the category of home runs) also walk comparably more. We will in fact find below that they are.⁴¹

Not only should hot hitters be walked more frequently, but the magnitude of this increase should match that which we find for better hitters. For example, suppose a hitter who hits home runs in 5 percent of his at bats is walked twice as much as one who hits home runs in only 2 percent of his at bats. If teams are forming proper inferences about the hot hand, it should not matter whether a player’s current 5% likelihood for hitting a home run is due to this being his long run average or whether he is normally a 3% home run hitter who is in a hot enough state that the best

⁴⁰In independent correspondence, one of the authors of this study, Mitchel Lichtman, described three additional hot hand studies he conducted. To our understanding, every one of the three had at least one of these two above mentioned drawbacks in the measurement of player ability.

⁴¹We speak throughout this section as if walks are entirely determined by the pitching (defensive) team. This is not a bad approximation: hitters can only be walked when the pitcher, either intentionally or unintentionally throws 4 pitches outside of the strike zone. However, there are offensive adjustments as well – for example, a hitter can be more (or less) discerning of the pitches at which he chooses to swing. And hitters may change their strategy depending on their current state. Hence, more precisely, we are measuring the joint endogenous response of pitchers and hitters, and whether this joint response corresponds the same way for hot hitters as for good hitters. With this change in terminology, all of our interpretations on hot hand inferences remain the same.

current inference is 5%. Hence, we ask whether this response is the same, and interpret differences to be indicative of the accuracy of inferences that opposing teams make with regard to hot-hands. Our findings indicate that opposing teams infer future performance from the past 25 at bats in a manner that indicates correct inferences about a hot hand, but they tend to “overreact” to very recent performance (i.e., performance in the last 5 at bats). Note that this comparison we make between good hitters and hot hitters is agnostic as to whether teams’ strategies on the degree to which they pitch around better players is in general optimal. Rather, we simply ask whether hitters with identical predicted outcomes are treated the same, whether the predicted outcome is from long-term ability or a hot hand.

5.1 Methodology

In order to conduct these tests, we re-estimate regressions of the format given in equation (1) making several necessary adjustments. Most importantly, to predict walks, we use measures of long-term ability and streakiness with respect to hits, extra-base hits, and home runs rather than using performance with respect to walks. In addition, we control for the batters ability using only information available to opposing teams at the time the at bat occurred (as opposed to using the windowed ability control described in Section 3.4). This adjustment is to reflect the fact that pitchers and managers do not have information about performance in future attempts when deciding how to pitch to the current batter. To compensate for the fact that using only previous attempts from the current season will result in a noisier measure of ability, we also control for the players ability using performance from the previous two seasons. Finally, we include controls for a number of situational factors (runners on base, score, number of outs, and lineup position) that affect the expected cost to the defensive team from walking a batter. We estimate predictability of home runs and extra base hits using the same control variables as used in the walks regressions.⁴²

Our test then consists of comparing the effect that long-term performance has on walks with the effect of short-term performance on walks, given the relative effect that these each have on outcomes of each of our other statistics. For example, in the case of home runs, we have estimates of how both long-term performance and short-term performance translate into the likelihood of hitting a home run and the likelihood of obtaining a walk through two regressions (one for home runs and one for walks). We then compare whether an increased likelihood in home runs due to a hot hand corresponds with the same increase in walks as does an equivalent increased likelihood in hitting a home run due to better long-term performance. This comparison involves a simple comparison of two ratios: 1) the ratio of regression coefficient for how short-term performance affects walks with the coefficient on how short-term performance affects home runs, and 2) the

⁴²For example, if the bases are loaded, then walking the current batter necessarily yields a run for the batting team and is therefore pitching around a batter in this situation is extremely costly to the defense. On the other hand, walking a batter when there is a lone runner on second base (so that after the walk there are runners on both first and second base) creates a force out situation on third base, which makes it considerably easier for the defense to get the lead runner out in the subsequent attempt.

same ratio of coefficients for long-term performance. If defenses form inferences accurately, the two ratios of regression coefficients should be equal. And assuming that defenses are relatively accurate in forming inferences from long-run ability, any deviation between such ratios can be interpreted as an inaccurate inference on the magnitude of the hot hand. Thus, if defenses overestimate (underestimate) the magnitude of the hot-hand effect, then the ratio of the coefficient on short-term performance should be higher (lower) than the ratio of the coefficients on long-term performance.

To test the null hypothesis, we use a non-linear Wald test comparing the ratio of the coefficient on short-term performance to the ratio of coefficients on long-term measures of performance (e.g., performance in the previous season). To compute the test statistic, we jointly estimate the covariance matrix of the coefficients across the two regression models clustering observations at the batter level. Under the null-hypothesis that the ratio across N measures are equal, the test statistic has a limiting chi-squared distribution with $N - 1$ degree of freedom.

5.2 Results

Table A.16 first confirms the well-known results that better hitters walk more frequently. That is, batters that hit more home runs and/or have more extra base hits in the long-run are walked more frequently. For example, column (1) illustrates that a batter who has hit home runs in 5% of all at bats in the current season (excluding the last 25 attempts) as well as the previous two seasons is $5\% \times (0.150 + 0.374 + 0.280) = 3.23$ percentage points more likely to get walked than a batter who has hit no home runs over the same time period. This corresponds to roughly a 30% increase in the likelihood of being walked. Not surprisingly, a higher ability in home runs and extra base hits leads to more walks than a higher batting average. Batting average is most frequently increased by single base hits, which are only slightly better for the offensive team than a walk, and therefore should not give rise to much defensive avoidance (i.e., pitching around the hitter). Defenses pitch around power hitters, especially those who hit home runs, much more frequently than around singles hitters.

Table A.16 also provides evidence that hot hitters also walk more frequently. In particular, we find that recent home runs by a batter are a strong predictor of whether the batter gets walked in the current at bat (columns (1) and (3)). Also, as with long-term performance, recent hits predict of walks (column (2)), with the effect primarily coming from extra base hits (doubles, triples and home runs) rather than singles (column (4)).

Table A.17 compares the predictability of home runs and extra base hits, from both long-run and short-run performance, to the change in walks. Columns (1) and (2) give estimates for the predictability of home runs and the likelihood of being walked based on the batter's home run per at bat ratio in: (i) the last 25 at bats (row 1), (ii) all at bats in the current season prior to the last 25 (row 2), (iii) the previous season (row 3) and (iv) the season two years ago (row 4). Column (3) then gives the ratio of the coefficient from column (2) divided by the coefficient from column (1). As noted above, if defenses correctly estimate the predictability of past performance, the ratio

of the coefficients across the four rows should be equal. And indeed we find that the ratios appear quite similar across the four rows ranging from 1.13 to 1.55 with no obvious pattern. While the ratio on the last 25 at bats is largest, we cannot reject the null hypothesis that the ratios are equal across all four rows even at a 5% confidence level. We perform the same test and obtain similar findings using extra base hits, by comparing the ratios of coefficients from column (5) to those in column (4). Once again, the null hypothesis, that the ratio of the coefficients in column (6) are the same across the four rows, cannot be rejected.⁴³

Hence, results in Table A.17 support the notion that opposing teams draw correct inferences about the hot hand of opposing batters using the last 25 at bats. The effect that this has on predicted outcome is treated in a statistically indistinguishable manner from how they treat better batters based on long-run performance. One might argue, however, that even if opposing teams form accurate inferences on the hot hand from the last 25 at bats, they might still be unduly influenced by the very most recent outcomes.

To test for this possibility, we break up divide recent performance into finer intervals and repeat the same regressions and tests. Table A.18 illustrates the results. The first 3 columns are similar to to columns (1)-(3) from Table A.17, except that recent performance in home runs is decomposed into three disjoint intervals of 10 at bats. Columns (4)-(6) further decomposes a batter's recent performance in home runs into five disjoint intervals of 5 at bats.

As before, under the null hypothesis that managers correctly estimate the predictability of past performance, the ratio of the coefficients in column (3) should be the same across the six rows. Here, however, we find that the most recent outcomes have a disproportionate impact on walks. Using intervals of 10 at bats, we find that the ratio on home run performance in the most recent 10 at bats (2.352) is significantly higher than the other five rows. This is confirmed by a non-linear hypothesis test, which strongly rejects the null that the first ratio is equal to any (or all) of the other five. This finding is even stronger when we use intervals of 5 at bats, as indicated in columns (4) to (6) of Table A.18. Column (4) shows that all else equal, a batter that has two home runs in his last five at bats (i.e., $hr_ab_1_5$ takes a value of 0.4) is about as likely to hit a home run in his next at bat (i.e., conditional on not being walked) as a batter who has hit two home runs in the five at bats before that (i.e., $hr_ab_6_10=0.4$), or the five at bats before that (i.e., $hr_ab_11_15=0.4$). However, Column (5) shows that a batter who has hit two home runs in his last five at bats is *significantly* more likely to be walked in his next plate appearance than a batter who hit two home runs in the five at bats before that, or the five before that. In fact, the coefficient on the most recent 5 attempts (3.944) is 3 to 4 times larger than any of the ratios and we can strongly reject the null hypothesis that the ratio on the most recent five at bats is equal to any (or all) of the seven other ratios. It is worth noting that when the last 5 at bats are decomposed further into the last two at bats and the three at bats prior, the ratio of the coefficients on *both* remains large, 4.980

⁴³For home runs, the test statistic under the null that all ratios are equal has a p-value of 0.0625. For extra base hits, the test statistic under the null hypothesis that all ratios are equal has a p-value of 0.2829.

and 2.969 respectively, and statistically different from each of the other seven ratios at confidence levels of 1% and 10% respectively.

Hence, while we find that opposing teams appear to draw correct inferences about the hot hand of opposing batters using the last 25 at bats (i.e., roughly the last week of performance), they tend to over react to streaky performances during the last 5 at bats (i.e., within the same game or the immediately prior game). Moreover, this overreaction is fairly large and strongest in the the last two attempts. This result is consistent with a version of the hot-hand fallacy in which significant streaky behavior exists and yet agents have a tendency to overestimate its magnitude - albeit, they only overestimate the importance of the very most recent outcomes, and seem to draw correct inferences on the hot hand for a somewhat longer history (which in turn has more power in predicting outcomes). On the other hand, these findings are in contrast with the predictions of the Gambler's Fallacy (Tversky and Khaneman, 1971). That is, if agents mistakenly believe in systematic reversals then one would expect them to under-react to very recent streaky performance (Rabin and Vayanos (2010)).

6 Simulations

In this section, we consider a simple Markov switching model of streakiness similar to those that have been used in previous work (see e.g., Stern and Morris (1993)). Our goal here is not to propose or test a specific model of streakiness; we remain agnostic to the specific form that streaky behavior may take. Rather, we present this simple model to give some intuitive sense of magnitudes of streaky behavior that are consistent with our empirical results. That is, we will ask what parameters in the Markov model would generate streakiness on the scale that we have observed, and we will in turn argue that these appear to be much in line with common intuitive notions of the hot hand in sports. In order to do so, we will use the Markov model to generate data, conduct regressions on simulated data, and compare our empirical estimates to those generated by the simulated data. As expected, given our imperfect empirical inference of the underlying state, we find the difference between performance in high and low states in the model is greater than the difference in performance based on inferences of the state from recent history. The magnitude of the difference is surprisingly large. In addition, the simulations demonstrate the need for a significant amount of data (on the order of 10^5 observations) in order to have the statistical power needed to identify streaky behavior.

6.1 A Simple Model of Streakiness

A player's ability follows a Markov-switching process. There are three underlying states: $\omega \in \{\text{Hot}(H), \text{Normal}(N), \text{Cold}(C)\}$. In the normal state, the player's (unconditional) probability of success is μ . When the player is hot (cold) the probability of success is $\mu + \Delta$ ($\mu - \Delta$). The transition matrix, M , is symmetric, with transition probabilities chosen such that the steady state distribution across states is $(\kappa, 1 - 2\kappa, \kappa)$ with the half life of the H and C state of $T_{1/2}$.

2016 Research Papers Competition
Presented by:

Note that Δ captures the magnitude of the streakiness effect while κ measures its prevalence.⁴⁴ These are the two key parameters of interest. We begin by simulating data for different values of these two parameters and estimating the least squares regression model given in (1) using R_{it}^L as the measure for the player's current state and the windowed ability given by (3).

For each pair of values (Δ, κ) , we simulate 100 data sets according to this process, each containing 1,000 players with 2,000 attempts per player (a total of two million observation). Motivated by the category of hits for batters (see Table A.2), the ability of each player, μ , is drawn from a distribution with mean 0.270 and standard deviation .025. The results are displayed in Table A.22, in which we report the mean estimate for γ and the mean t-statistic across the 100 simulations. In Table A.23, we perform the same exercise where batters are assumed to be homogenous (each with $\mu = 0.270$). For simplicity, we assume that all other parameters are uniform across all players and fix $T_{1/2} = 25$, which corresponds to the length of the history used to evaluate the players state.

Both tables yield similar results. First, as expected, higher Δ generate larger estimates for the coefficient and its statistical significance. Increasing κ has a similar effect; the more frequently a player is hot (cold), the more likely it is that strong (weak) recent performance is indicative of being hot (cold).

The most notable observation comes from a comparison to the empirical estimates from the baseball data. The analogous estimates to Tables A.22 and A.23 are given in column (1) of Tables A.5 and A.6, where we estimated a coefficient of 0.0308 for batters and 0.0448 for pitchers. Notice these estimates are consistent with a Δ between 0.04 and 0.05 (40-50 batting points) and κ between 5 and 10%.⁴⁵ This corresponds to roughly a 20% increase in the likelihood of getting a hit for a hot player compared to an average player and a 40% increase relative to a cold player. Note that this magnitude for Δ is 5 to 7 times larger than the difference in performance between players that we infer to be hot or cold in our dummy specifications of Tables A.5 and A.6 (columns (4)-(5)). That is, if the Markov model was the true underlying model, our dummy specifications are sufficiently noisy in identifying true hot and cold players, that they only yield a fraction of the underlying difference in ability based on state.⁴⁶

Thus, while we found roughly $\frac{1}{2}$ to 1 standard deviation difference between a player empirically identified as hot and cold player in the previous section, in order to generate this result from our Markov model we would need the true difference between correctly identified hot and cold players to be on the order of 2.5 to 3 standard deviations (roughly 80 batting points). GVT surveyed fans on their perceptions of a hot hand, and argued that they were implausibly large, even if a small effect existed. However, these responses, which were on the order of 5-10 percentage points, are

⁴⁴When either $\Delta = 0$ or $\kappa = 0$, the data generating process is i.i.d.

⁴⁵Similar conclusions obtain when estimating alternative specifications. Results available upon request.

⁴⁶Of course, if we were confident that the Markov model was the correct model, we could estimate who is hot and cold with a more efficient procedure. That is not our intent here, as we do not want to test one particular model of streakiness. Rather our point is that true underlying streakiness will be significantly higher than differences in performance from empirically inferred streakiness based on recent performance, due to both noise and model misspecification.

very much in line with the differences implied by our simulations; if respondents to the GVT survey believed they were being asked about the difference in performance between a correctly identified hot player and a neutral player, without adjusting defenses (which to us is the most natural reading of the question that GVT asked), then they gave responses much in line with the parameters of the Markov model needed to generate our empirical estimates.

6.2 Power

We have argued above that in sports where defenses can adjust freely, one should not expect to find short-run predictability in player performance, whether or not a hot-hand effect exists. In contrast, this should not be the case in sports in which defensive adjustments are sufficiently costly. It is notable then that tests in a number of other sports have not found evidence of a hot hand. However, many of these tests suffer from a second drawback distinct from endogeneity, that being the lack of power needed to reject their null hypothesis. In Tables A.24-A.25, we use the Markov model to show the number of observations needed to detect streakiness using our methodology is on the order of 10^5 , while the previous studies that we are aware of have used data in which the number of observations is *at least* one order of magnitude lower. For example, if the true model has the parameters $\kappa = 0.05$ and $\Delta = 0.05$ (i.e., a 10% difference between hot and cold) then with 10,000 observations our tests would reject the null hypothesis (of $\gamma = 0$) at a 1 percent (5 percent) confidence level in only 4% (11%) of the 100 simulations. With 100,000 observations, the rejection rates increase to 22% and 48%. Moreover, tests employed elsewhere in the literature, with shorter recent histories and no panel, have lower power, and would require even more data than our tests do (see e.g., Wardrop (1999) for further discussion).

7 Discussion

In this section we discuss two topics: First, we address endogeneity in other sports, and hypothesize that one will be likely to find streakiness in a given sport if and only if endogenous responses do not equate margins across players. Furthermore, we argue that this latter condition is likely to be linked to the variation in long-run ability. Second, we discuss the magnitude of the effects we find, and argue that at a minimum, they appear consistent with conventional notions of hot hands.

7.1 Other Sports: Hot Hands and Endogeneity in General

We focus in this paper on the distinction between basketball (for which streakiness has been extensively tested) and baseball (which we test). But as noted in Section 1, the distinction between sports which do and do not have an equating endogenous response motivates broader questions on the existence of a hot hand in other sports and activities. We hypothesize that skilled activity will generally exhibit streakiness – that is, there will be transitory components to ability as well as

long-run components. However, we would expect to find evidence of such a hot hand in outcomes (i.e., streakiness in outcomes) if and only if the activity at hand does not permit an endogenous response that is likely to equate margins, as in basketball.

To this end, we would predict that golf, bowling, track and field, and shooting (all of which admit no defense), and baseball, boxing, and chess (which have limited or no scope for transferring defensive resources across players), should all exhibit streakiness in outcomes. Indeed, golf and bowling are two sports in which there is some direct evidence of a hot hand (see Dorsey-Palmateer and Smith, 2004 and Arkes, 2014). Conversely, soccer, hockey, and basketball, all of which involve fluid defenses that choose how to allocate resources across players, as well as offenses that choose how to allocate scoring opportunities across players, should exhibit limited or no evidence for streakiness in shooting/scoring probabilities. One obvious category of exception in soccer (hockey) is in penalty kicks (a shootout) where the goalie faces each offensive player sequentially.

One difficulty is that while in some of these cases the distinction between the presence and absence of a defensive response that equates margins should be obvious (there is no defense in golf or bowling), in other cases it is more subtle and depends on the details of the sport (the distinction between basketball and baseball). Here, however, we would argue that there is a simple intuitive way to identify the presence of such a defensive response, and consequently, whether one would expect to find streakiness in outcomes.

In particular, endogenous responses, when available, should equate margins both across permanent and transitory differences in abilities. Thus, we hypothesize that a straightforward manner to identify settings where available endogenous responses are sufficient to negate short-term streakiness is to identify settings with permanent differences in outcomes. If player performance in a particular statistic is correlated with other measures of player quality (such as player salary, or playing time, or awards), and if it is persistent across long stretches of time (i.e., over years), that is indicative of a setting where endogenous responses are not sufficient to equate margins, and hence where we would also expect to find short-run streakiness. In contrast, if instead performance does not correlate with other measures of ability, and if performance does not persist across years, this is indicative of the presence of an endogenous response which equates margins, and we would not expect to find short-run streakiness. This is line with the distinction between basketball shooting and baseball hitting that Figures 1 and 2 in the introduction highlight. Similarly, baseball all-stars typically have higher batting averages and higher home run percentages and lower earned run averages than average players, but basketball all-stars do not on average have higher shooting percentages than average players.⁴⁷

⁴⁷One might expect to see some small correlation between performance and other measures of quality even when there is an endogenous response, if the endogenous response is significant but incomplete in some settings (i.e. inframarginal shots in basketball), or if the statistic does not perfectly measure the performance over which the defense is optimizing (i.e., equating expected shot value in basketball is not the same as equating shooting percentages). But these correlations should be much smaller than for other settings where there is either a weak or no endogenous response at all.

Indeed, the difference between field goal shooting and free-throw shooting in basketball is particularly revealing here. As we have already noted, there is no discernable difference between the field goal percentage of the players considered the best offensive players in basketball and average players. In sharp contrast, there is a very significant difference in free-throw shooting (which is not defended and which does not permit an endogenous response) between players, with players who are considered better outside shooters predictably and persistently shooting far better in free-throws than those who are not considered good outside shooters.

In summary, we hypothesize that hot hands are natural in skilled activities – skill is likely to vary both in the short-run and the long-run. Streakiness from short-run variation in skill, however, will only be observed in outcomes if there is not an endogenous equilibrium response which equates margins. One straightforward way in sports to identify whether such a response likely exists is to observe long-run performance, as such a response should apply to both long-run and short-run observable components of skill. Consequently, we hypothesize that the presence or absence of long-run ability in outcome data in a sport will predict the presence or absence of short-run streakiness in the same statistic. We should find short-run predictability in sports statistics if those same statistics correlate with other long-run measures of players ability, and if player performance in those statistics is persistent across long periods.

7.2 Magnitude of the Hot Hand

The literature has argued that even if players do exhibit a hot hand, common perceptions overestimate the magnitude of this effect (e.g., see Rabin (2002)).⁴⁸ We have presented evidence above that both supports and challenges this view. On one hand, we have shown that the hot hand is rather significant and strategically relevant in baseball. A hot (cold) median ability player will look more like a 75th (25th) percentile player. Such magnitudes are very much in line with managerial playing decisions, whereby managers will almost always start their best players, but will frequently switch between marginal starters and nonstarters based on recent performance and perceived streakiness.⁴⁹ Moreover, using walks to measure how opposing teams react to hot and cold hitters, we find that these reactions are consistent with teams making accurate inferences on the magnitude of the hot hand as predicted from the last 25 at bats. That is, defenses do react, and react in a manner and magnitude roughly consistent with how they react to an equivalent difference in long-term ability.

⁴⁸There are also a number of surveys, dating back to GVT that make this point. We would argue, however, that the survey questions that attempts to measure beliefs on hot hands are often ambiguous and subject to different interpretations than those given by their authors. For instance, when GVT ask about such beliefs, it is not clear whether the responder is likely to interpret the question to be about probabilities conditioning on an actual hot state or instead to be about probabilities conditioning on the last outcome (regardless of state), and whether the conditioning should be with or without taking defensive adjustments into account. As we discuss in Section 6, probabilities conditional on an actual hot state – which seems to us to be a natural manner for respondents to interpret the GVT survey question – should be significantly larger than than probabilities conditional on making the last shot to which GVT compare their survey results.

⁴⁹And it is notable that coaches routinely refer to recent performance when justifying such line-up moves, saying things like a certain player is “in a groove”, is “swinging a hot bat”, or has been “throwing very well lately”.

On the other hand, we have also shown that teams significantly overreact to the last 5 at bats. Thus defenses appear to overstate the importance of the very most recent performance in predicting opponents' outcomes, while adjusting correctly for the hot hand over longer (but still short-term) frequencies. We believe that our paper is novel in testing such reactions and in finding this distinction.

It is plausible that a player or a coach would be able to better identify when they are in a hot or cold streak better than the econometrician who only observes outcomes. Hot and cold streaks are likely to manifest in other physical or psychological displays such as confidence, harder pitches, minor injuries, etc. observed by the player/coach but not observed by the econometrician. Players and coaches can also observe when a hit was a solid hit, and when it was a lucky hit. This better identification of states would in turn magnify the difference in performance between times identified as hot and cold. In the Markov model we presented above in Section 6, the performance difference between actual states was roughly 5 times larger than the performances difference based on observing hot or cold recent outcomes. To the extent that players and coaches can do a better job identifying the actual state with their additional information, one would expect to see more active strategic responses to their perceptions of player hotness.

8 Conclusion

We test for a hot hand (i.e., short-term streakiness in performance) in Major League Baseball using panel data. We consider ten statistical categories, and find strong evidence of a hot hand in all of them. Moreover, the magnitudes are significant; strong recent performance relative to being in a neutral state corresponds to roughly a one quartile increase in the distribution of present performance. Thus, a 50th percentile hitter will hit like a 75th percentile hitter following strong recent performance.

Our results are in notable contrast to the majority of the sports hot-hand literature, which has mainly found either little evidence for a hot hand, or evidence for only a very weak hot hand, often employing basketball shooting data. We argue that a primary explanation for this difference is endogenous defensive responses: basketball presents sufficient opportunity for defensive responses to equate shooting probabilities across players whereas baseball does not. As such, much prior evidence on the absence of a hot hand despite widespread belief in its presence should not necessarily be interpreted as a cognitive mistake, but rather, as an endogenous defensive response.

We additionally formulate a test for endogenous responses using walks, which in turn allows us to test whether the large hot hand that we find is in line with inferences that teams have regarding the hot hand. Here our results are mixed: We find that teams respond to hot opposing batters in a way that is consistent with drawing the correct inference about the magnitude of the hot hand over an opponent's last 25 at bats. However, at the same time, our tests also indicate that opposing teams over-react to very recent performance from the last 5 at bats. This last result is in line with

a hot-hand fallacy; albeit one that is more qualified and limited than what is frequently portrayed.

We also provide a simple heuristic for identifying a priori which sports are likely to permit an equating endogenous response and discuss potential implications for identifying the hot-hand effect or ability in other settings. Defenses should treat permanent and transient skill differences of opposing players in a similar manner: defenses are likely to adjust to equate margins across hot and cold players if and only if they also adjust to equate margins across good and bad players. Hence we should expect to see streakiness in a statistic if and only if we also see permanent quality differences across players in the same statistic.

References

- Robert M Adams. The 'hot hand' revisited: Successful basketball shooting as a function of intershot interval. *Perceptual and Motor Skills*, 74(3):934–934, 1992. URL <http://www.amsciepub.com/doi/pdf/10.2466/pms.1992.74.3.934>.
- Robert M Adams. Momentum in the performance of professional tournament pocket billiards players. *International Journal of Sport Psychology*, 1995. URL <http://psycnet.apa.org/psycinfo/1996-03492-010>.
- Gil Aharoni and Oded H Sarig. Hot hands and equilibrium. *Applied Economics*, 44(18):2309–2320, 2012. URL <http://www.tandfonline.com/doi/abs/10.1080/00036846.2011.564141>.
- Jim Albert and Jay Bennett. *Curve ball: Baseball, statistics, and the role of chance in the game*. Springer, 2003.
- S Christian Albright. A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association*, 88(424):1175–1183, 1993. URL <http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476395>.
- Jeremy Arkes. Revisiting the hot hand theory with free throw data in a multivariate framework. *Journal of Quantitative Analysis in Sports*, 6(1), 2010.
- Jeremy Arkes. Misses in "hot hand" research. *Journal of Sports Economics*, 14(4):401–410, 2013.
- Jeremy Arkes. The hot hand in golf. Working Paper, 2014.
- Michael Bar-Eli, Simcha Avugos, and Markus Raab. Twenty years of "hot hand" research: Review and critique. *Psychology of Sports and Exercise*, 7(6):525–553, 2006.
- Andrew Bocskocsky, John Ezekowitz, and Carolyn S. M. Stein. Heat check: New evidence on the hot hand in basketball. Working Paper, 2014.
- Daniel Chen, Tobias J. Moskowitz, and Kelly Shue. Decision-making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. Working Paper, 2014.
- Reid Dorsey-Palmateer and Gary Smith. Bowlers' hot hands. *The American Statistician*, 58(1):38–45, 2004. URL <http://www.tandfonline.com/doi/abs/10.1198/0003130042809>.
- Ronald Forthofer. Streak shooter?the sequel. *Chance*, 4(2):46–48, 1991.
- D Frame, E Hughson, and JC Leach. Runs, regimes, and rationality: The hot hand strikes back. Technical report, Working paper, 2003.
- David L Gilden and Stephanie Gray Wilson. Streaks in skilled performance. *Psychonomic Bulletin & Review*, 2(2):260–265, 1995. URL <http://link.springer.com/article/10.3758/BF03210967>.
- Thomas Gilovich, Robert Vallone, and Amos Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314, 1985.

- Stephen Jay Gould. The streak of streaks. *Chance*, 2(2):10–16, 1989. URL <http://www.tandfonline.com/doi/abs/10.1080/09332480.1989.10554932>.
- John Huizinga and Sandy Weil. Hot hand or hot head: The truth about heat checks in the nba. 2009.
- Franc JGM Klaassen and Jan R Magnus. Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96(454):500–509, 2001. URL <http://www.tandfonline.com/doi/abs/10.1198/016214501753168217>.
- Jonathan Koehler and Caryn Conley. The 'hot hand' myth in professional basketball. *Journal of Sport & Exercise Psychology*, 25:253, 2003. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1469609.
- Patrick D Larkey, Richard A Smith, and Joseph B Kadane. It's okay to believe in the "hot hand". *Chance*, 2(4):22–30, 1989. URL <http://www.tandfonline.com/doi/pdf/10.1080/09332480.1989.10554950>.
- Joshua Benjamin Miller and Adam Sanjurjo. A cold shower for the hot hand fallacy. *IGIER Working Paper*, 2014.
- Hiroto Miyoshi. Is the "hot-hand" phenomenon a misperception of random events? *Japanese Psychological Research*, 42:128–133, 2000. ISSN 0021-5368. doi: 10.1111/1468-5884.00138. URL <http://doi.wiley.com/10.1111/1468-5884.00138>.
- Matthew Rabin. Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117:775–816, 2002. ISSN 0033-5533.
- Matthew Rabin and Dimitri Vayanos. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, 77:730–778, 2010.
- Gary Smith. Horseshoe pitcher's hot hands. *Psychonomic bulletin & review*, 10(3):753–758, 2003. URL <http://link.springer.com/article/10.3758/BF03196542>.
- Hal S Stern and Carl N Morris. A statistical analysis of hitting streaks in baseball: Comment. *Journal of the American Statistical Association*, 88(424):1189–1194, 1993. URL <http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1993.10476397>.
- Daniel F. Stone. Measurement error and the hot hand. *The American Statistician*, 66(1):61–66, 2012.
- Seymour Swioff. *The Elias Baseball Analyst, 1988*. MacMillan Publishing Company, 1988.
- Tom Tango, Mitchel Lichtman, and Andrew Dolphin. *The Book: Playing the Statistical Percentages in Baseball*. Potomac Books, 2007.
- Amos Tversky and Daniel Kahneman. Belief in the law of small numbers. *Psychological Bulletin*, 76(2):105–110, 1971.
- Roger C Vergin. Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior*, 2000. URL <http://psycnet.apa.org/psycinfo/2000-15656-008>.

Robert L Wardrop. Simpson's paradox and the hot hand in basketball. *The American Statistician*, 49(1): 24-28, 1995. URL <http://amstat.tandfonline.com/doi/abs/10.1080/00031305.1995.10476107>.

Robert L Wardrop. Statistical tests for the hot-hand in basketball in a controlled setting. *American Statistician*, 1:1-20, 1999.

Jeffrey Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2 edition, 2010.



A Tables

A.1 Summary Statistics

variable	mean	sd	p25	p50	p75	p99	N
Batting	.273	.0292	.253	.272	.292	.342	3265
Homerun	.0332	.0195	.0185	.0308	.046	.0856	3265
Strikeout	.183	.0617	.138	.177	.222	.347	3265
OnBase	.345	.0377	.32	.343	.368	.444	3265
Walk	.0994	.0352	.0749	.0945	.12	.193	3265
PlateAppearances	492	120	388	498	596	702	3265
AtBats	442	107	351	449	533	640	3265

Source:

Table A.1 – Batter-year Summary Statistics. This table provides summary statistics of MLB batting statistics at the batter-year level over the duration of the sample period (2000-2011). Batter-years in which the batter had fewer than 300 plate appearances have been excluded from this table.

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

Source:

Table A.2 – Batter Summary Statistics. This table provides summary statistics of MLB batting statistics at the batter level over the duration of the sample period (2000-2011). Batter-years in which the batter had fewer than 300 plate appearances have been excluded from this table.

variable	mean	sd	p25	p50	p75	p99	N
Batting	.264	.0309	.245	.266	.285	.336	2607
Homerun	.0304	.0107	.0232	.0298	.0369	.06	2607
Strikeout	.191	.0554	.151	.183	.223	.352	2607
OnBase	.333	.0325	.312	.333	.354	.406	2607
Walk	.093	.0263	.0744	.091	.109	.162	2607
PlateAppearances	570	219	343	560	783	947	2607
AtBats	519	204	309	505	717	885	2607

Source:

Table A.3 – Pitcher-year Summary Statistics. This table provides summary statistics of MLB pitching at the pitcher-year level over the duration of the sample period (2000-2011). Pitcher-years in which the pitchers faced fewer than 300 batters have been excluded from this table.

variable	mean	sd	p25	p50	p75	p99	N
Batting	.264	.028	.247	.268	.283	.327	820
Homerun	.0307	.00985	.0249	.0304	.0359	.0571	820
Strikeout	.191	.0538	.153	.182	.223	.345	820
OnBase	.335	.0284	.319	.336	.353	.401	820
Walk	.0965	.0232	.0803	.0948	.11	.158	820
PlateAppearances	1814	1978	432	979	2370	8624	820
AtBats	1650	1818	391	879	2161	7893	820

Source:

Table A.4 – Pitcher Summary Statistics. This table provides summary statistics of MLB pitching at the pitcher level over the duration of the sample period (2000-2011). Pitcher-years in which the pitchers faced fewer than 300 batters have been excluded from this table.



A.2 Main Results

	(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

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

Table A.5 – Dependent Variable: Hits (Batters). This table reports estimates for the batter regressions where the dependent variable indicates a hit. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the batter’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the batter’s state: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the batter’s state use his 25 most recent at bats ($L=25$). Note that all plate appearances that did not constitute an at bat were excluded from the regressions. Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	Dist5
main						
state	0.0448*** (9.06)	0.230*** (9.03)	0.0112*** (7.15)			
hot				0.00748*** (3.89)	0.00717*** (3.83)	0.00412* (2.49)
cold				-0.00787*** (-4.31)	-0.00815*** (-4.19)	-0.00861*** (-4.25)
pitcher_ability	0.380*** (16.18)	1.954*** (15.96)	0.456*** (17.82)	0.408*** (16.95)	0.409*** (16.93)	0.425*** (16.82)
batter_ability	0.561*** (50.09)	2.940*** (48.57)	0.560*** (47.76)	0.561*** (50.14)	0.561*** (50.15)	0.560*** (48.13)
samehand	-0.0131*** (-10.44)	-0.0672*** (-10.43)	-0.0126*** (-9.66)	-0.0130*** (-10.39)	-0.0130*** (-10.39)	-0.0127*** (-9.84)
pitcher_home	-0.00866*** (-9.78)	-0.0447*** (-9.87)	-0.00810*** (-8.78)	-0.00877*** (-9.85)	-0.00877*** (-9.85)	-0.00829*** (-9.00)
Observations	1128156	1128156	1007607	1128156	1128156	1037519

t statistics in parentheses

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

Table A.6 – Dependent Variable: Hits (Pitchers). This table reports estimates for the pitcher regressions where the dependent variable indicates a hit. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the pitcher’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the pitcher’s state: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the pitcher’s state use his 25 most recent at bats ($L=25$). Note that all plate appearances that did not constitute an at bat were excluded from the regressions. Standard errors are clustered at the pitcher level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	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

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

Table A.7 – Dependent Variable: On Base (Batters). This table reports estimates for the batter regressions where the dependent variable indicates the batter reached base. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the batter’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the batter’s State: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the batter’s state use his 25 most recent plate appearances ($L=25$). Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	Dist5
main						
state	0.0696*** (15.54)	0.316*** (15.45)	0.0195*** (12.20)			
hot				0.0142*** (7.76)	0.0126*** (6.96)	0.0105*** (6.91)
cold				-0.0157*** (-8.65)	-0.0131*** (-6.96)	-0.00985*** (-5.20)
pitcher_ability	0.423*** (18.87)	1.921*** (18.52)	0.510*** (21.10)	0.468*** (20.06)	0.470*** (20.07)	0.467*** (18.79)
batter_ability	0.621*** (63.66)	2.845*** (63.17)	0.621*** (60.59)	0.621*** (63.74)	0.621*** (63.75)	0.621*** (61.33)
samehand	-0.0221*** (-17.33)	-0.100*** (-17.27)	-0.0217*** (-16.18)	-0.0220*** (-17.22)	-0.0220*** (-17.22)	-0.0219*** (-16.52)
pitcher_home	-0.0116*** (-13.94)	-0.0528*** (-14.03)	-0.0109*** (-12.38)	-0.0118*** (-14.06)	-0.0118*** (-14.08)	-0.0112*** (-12.78)
Observations	1344753	1344753	1196797	1344753	1344753	1242357

t statistics in parentheses

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

Table A.8 – Dependent Variable: On Base (Pitchers). This table reports estimates for the pitcher regressions where the dependent variable indicates the batter reached base. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the pitcher’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the pitcher’s State: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the pitcher’s current state (hot, cold or normal). In all columns, the measure of the pitcher’s state use his 25 most recent plate appearances ($L=25$). Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	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

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

Table A.9 – Dependent Variable: Home runs (Batters). This table reports estimates for the batter regressions where the dependent variable indicates a home run. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the batter’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the batter’s state: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the batter’s state use his 25 most recent at bats ($L=25$). Note that all plate appearances that did not constitute an at bat were excluded from the regressions. Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	Dist5
main						
state	0.0326*** (6.53)	1.032*** (6.48)	0.00366*** (4.62)			
hot				0.00340*** (4.02)	0.00341*** (3.93)	0.00327*** (4.59)
pitcher_ability	0.267*** (11.49)	8.691*** (11.68)	0.320*** (12.33)	0.291*** (12.24)	0.279*** (11.83)	0.291*** (11.62)
batter_ability	0.717*** (66.29)	21.32*** (77.54)	0.707*** (60.89)	0.717*** (66.32)	0.717*** (66.32)	0.709*** (61.86)
samehand	-0.00395*** (-8.56)	-0.124*** (-7.91)	-0.00381*** (-7.92)	-0.00394*** (-8.53)	-0.00395*** (-8.54)	-0.00379*** (-8.02)
pitcher_home	-0.00152*** (-4.48)	-0.0472*** (-4.19)	-0.00156*** (-4.52)	-0.00153*** (-4.48)	-0.00153*** (-4.50)	-0.00147*** (-4.30)
Observations	1128156	1128156	1007910	1128156	1128156	1037519

t statistics in parentheses

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

Table A.10 – Dependent Variable: Home runs (Pitchers). This table reports estimates for the pitcher regressions where the dependent variable indicates a home run. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the pitcher’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the pitcher’s state: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the pitcher’s state use his 25 most recent at bats ($L=25$). Note that all plate appearances that did not constitute an at bat were excluded from the regressions. Standard errors are clustered at the pitcher level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	Dist5
main						
state	0.0861*** (16.60)	0.578*** (16.80)	0.0194*** (13.79)			
hot				0.0111*** (6.98)	0.0144*** (7.71)	0.0120*** (7.85)
cold				-0.0116*** (-7.39)	-0.0172*** (-9.25)	-0.0126*** (-7.08)
batter_ability	0.793*** (81.98)	5.268*** (60.25)	0.889*** (103.56)	0.870*** (95.94)	0.876*** (97.92)	0.880*** (98.91)
pitcher_ability	0.820*** (64.87)	5.320*** (98.06)	0.815*** (62.07)	0.822*** (64.85)	0.821*** (64.90)	0.817*** (62.35)
samehand	0.0177*** (17.88)	0.126*** (17.81)	0.0176*** (17.16)	0.0177*** (17.71)	0.0177*** (17.75)	0.0177*** (17.17)
batter_home	-0.00824*** (-11.34)	-0.0580*** (-11.35)	-0.00863*** (-11.64)	-0.00828*** (-11.37)	-0.00825*** (-11.33)	-0.00856*** (-11.48)
Observations	1192266	1192266	1101941	1192266	1192266	1113087

t statistics in parentheses

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

Table A.11 – Dependent Variable: Strikeouts (Batters). This table reports estimates for the batter regressions where the dependent variable indicates a strikeout. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the batter's state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the batter's state: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter's current state (hot, cold or normal). In all columns, the measure of the batter's state use his 25 most recent at bats ($L=25$). Note that all plate appearances that did not constitute an at bat were excluded from the regressions. Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	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

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

Table A.12 – Dependent Variable: Strikeouts (Pitchers). This table reports estimates for the pitcher regressions where the dependent variable indicates a strikeout. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the pitcher’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the pitcher’s state: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the pitcher’s state use his 25 most recent at bats ($L=25$). Note that all plate appearances that did not constitute an at bat were excluded from the regressions. Standard errors are clustered at the pitcher level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	Dist5
main						
state	0.113*** (24.13)	1.133*** (23.78)	0.0209*** (18.60)			
hot				0.0157*** (12.46)	0.0187*** (13.72)	0.0138*** (11.40)
cold					-0.0146*** (-10.96)	-0.0125*** (-7.96)
batter_ability	0.683*** (37.90)	6.851*** (36.29)	0.809*** (42.20)	0.789*** (39.63)	0.788*** (40.03)	0.791*** (39.09)
pitcher_ability	0.790*** (60.53)	8.793*** (67.32)	0.789*** (57.65)	0.794*** (60.71)	0.793*** (60.71)	0.792*** (58.22)
samehand	-0.0162*** (-19.10)	-0.186*** (-19.60)	-0.0165*** (-18.66)	-0.0163*** (-18.92)	-0.0163*** (-18.98)	-0.0165*** (-18.45)
batter_home	0.00674*** (13.56)	0.0757*** (13.60)	0.00694*** (13.42)	0.00687*** (13.60)	0.00680*** (13.56)	0.00696*** (13.43)
Observations	1458194	1458194	1352471	1458194	1458194	1368859

t statistics in parentheses

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

Table A.13 – Dependent Variable: Walks (Batters) This table reports estimates for the batter regressions where the dependent variable indicates a walk. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the batter’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the batter’s State: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the batter’s current state (hot, cold or normal). In all columns, the measure of the batter’s state use his 25 most recent plate appearances ($L=25$). Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Logit	OLS_dist	Prop5	Add5	Dist5
main						
state	0.0436*** (8.46)	0.532*** (8.62)	0.0749*** (5.37)			
hot				0.00651*** (5.46)	0.00552*** (4.50)	0.00330*** (3.68)
cold					-0.00559*** (-4.11)	-0.00392* (-1.99)
pitcher_ability	0.664*** (44.48)	7.881*** (35.96)	8.817*** (37.50)	0.702*** (46.01)	0.705*** (46.01)	0.715*** (45.76)
batter_ability	0.629*** (67.39)	7.190*** (78.46)	7.193*** (74.66)	0.629*** (67.39)	0.629*** (67.41)	0.624*** (64.83)
samehand	-0.0137*** (-18.24)	-0.171*** (-17.94)	-0.174*** (-17.02)	-0.0137*** (-18.21)	-0.0137*** (-18.21)	-0.0138*** (-17.57)
pitcher_home	-0.00535*** (-10.73)	-0.0680*** (-11.07)	-0.0618*** (-9.59)	-0.00543*** (-10.77)	-0.00542*** (-10.79)	-0.00509*** (-9.74)
Observations	1344753	1344753	1201352	1344753	1344753	1242357

t statistics in parentheses

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

Table A.14 – Dependent Variable: Walks (Pitchers). This table reports estimates for the pitcher regressions where the dependent variable indicates a walk. Column (1) gives the estimates from the OLS regression given in equation (1) and using the baseline measure of the pitcher’s state as estimated by equation (3). Column (2) gives estimates for the logistic regression given in equation (2) using the same independent variables. Columns (3)-(6) contain OLS estimates using alternative measures of the pitcher’s State: column (3) uses the estimate of state given in equation (5); columns (4)-(6) use the additive, proportional and distribution based cutoffs respectively to measure the pitcher’s current state (hot, cold or normal). In all columns, the measure of the pitcher’s state use his 25 most recent plate appearances ($L=25$). Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

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

Table A.15 – Normalized Magnitude of the Hot Hand Effect. This table summarizes the magnitude of the findings from Tables A.5-A.14 using the additive cutoffs model (column (5)).

Mean – The mean of the distribution across players of the statistic

Std Dev – The standard deviation of the distribution across players of the statistic

Hot - Cold – The difference in the coefficients for a recent hot and cold history, as given by our additive cut-off specification (column (5) in our main results tables)

(Hot - Cold)/Mean – (Hot - Cold) divided by Mean

(Hot - Cold)/Std Dev – (Hot - Cold) divided by Std Dev

[†] For home runs, our cut-off specification does not specify a cold 5% outcome, since for over 40% of the observations, there are no home runs hit in the last 25 at bats. In order to make our number comparable here to the other statistics, the number reported here is twice the hot coefficient. (Note that for all other statistics, the hot and cold coefficients are of similar magnitude and opposite sign.)

	(1)	(2)	(3)	(4)	(5)
	Walk	Walk	Walk	Walk	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

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

Table A.16 – Endogenous Response: Walks (Batters). This table estimates the endogenous response of the defense to the batter’s ability in hits, home runs and extra base hits. The dependent variable indicates a walk. Column (1) provides estimates for the endogenous response to batter’s State (as estimated by equation (3)) and ability in home runs (as controlled for by performance this season excluding row 1, as well as performance in the prior to seasons). Column (2) and (3) provides estimates for similar regressions using batter’s State and ability in extra base hits and all hits respectively. Column (4) combines home runs and hits and Column (5) combines extra base hits with hits. Controls for pitcher ability, the platoon effect, whether the batter is in his home stadium, number of outs and runners on base are included in the regression but omitted from the table for parsimony. All regressions also include stadium \times year dummy control variables.

	(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

Table A.17 – Predictability and Endogenous Response to Home Runs and Extra Base Hits. This table compares the predictability of home runs (column (1)) and extra base hits (column (4)) to the likelihood of being walked (columns (2) and (5)). Column (3) reports the ratio of the coefficients on the regressors in column (2) to those in column (1). Column (6) reports the ratio of the coefficients on the regressors in column (5) to those in column (4). The number in parentheses in columns (3) and (6) is the test statistic from the Wald test from the pairwise test that the ratio of the coefficients on the performance measure in this row is equal to the ratio of coefficients on the short-run performance measure. Controls for pitcher ability, the platoon effect, whether the batter is in his home stadium, number of outs and runners on base are included in the regression but omitted from the table for parsimony. All regressions also include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	Homerun	Walk	Ratio	Homerun	Walk	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

Table A.18 – Predictability and Endogenous Response to Home Runs at Finer Performance Intervals.

This table compares the predictability of home runs (columns (1) and (4)) to the likelihood of being walked (columns (2) and (5)). Column (3) reports the ratio of the coefficients on the regressors in column (2) to those in column (1). Column (6) reports the ratio of the coefficients on the regressors in column (5) to those in column (4). The number in parenthesis in columns (3) and (6) is the test statistic from the Wald test from the pairwise test that the ratio of the coefficients on the performance measure in this row is equal to the ratio of coefficients on the short-run performance measure. Controls for pitcher ability, the platoon effect, whether the batter is in his home stadium, number of outs and runners on base are included in the regression but omitted from the table for parsimony. All regressions also include stadium \times year dummy control variables.



A.3 Learning Tests

Statistic	LTObs	STObs	LTCoeff	STCoeff	STNull	Learn	Fstat	Prob>F
Hits Bat	402	25	0.3660	0.0308	0.0228	0.7390	2.36	0.125
On Base Bat	448	25	0.5390	0.0640	0.0301	0.4700	51.78	0.000
Home Runs Bat	402	25	0.7020	0.0749	.0437	0.5829	38.35	0.000
Strike Outs Bat	402	25	0.7930	0.0861	.0493	0.5728	44.90	0.000
Walks Bat	448	25	0.6830	0.1130	0.0381	0.3373	247.57	0.000
Hits Pitch	551	25	0.3800	0.0448	0.0172	0.3849	29.61	0.000
On Base Pitch	593	25	0.4230	0.0696	0.0178	0.2562	127.10	0.000
Home Runs Pitch	551	25	0.2670	0.0326	0.0121	0.3716	15.45	0.000
Strike Outs Pitch	551	25	0.7200	0.1180	0.0327	0.2768	230.60	0.000
Walks Pitch	593	25	0.6640	0.0436	0.0280	0.6420	8.64	0.003
Average Bat						0.5404		
Average Pitch						0.3863		
Overall Average						0.4633		

Table A.19 – Learning vs. Streakiness. Linear Learning. LRObs gives the average number of observations per player for the ability measure. STObs is the number of observations in the recent history (always 25). LTCoeff is the coefficient on player ability in our base regression. STCoeff is the coefficient on the short-run history in our base regression. ST Null is the anticipated coefficient on the short-run history under the null of learning, but no streakiness. Learn gives the proportion of STNull to the realized STCoeff; i.e., the fraction of the coefficient on the short-run history that is attributable to learning. The second to last column is the F-statistic from the test from a Wald test as to whether the estimated coefficient can be explained by the null hypothesis of learning absent streakiness and the last column is the p-value.

Statistic	STCcoeff	STCcoeffOL	Streak	Learn
Hits Bat	0.0326	0.0053	0.1632	0.8368
On Base Bat	0.0628	0.0279	0.4443	0.5557
Home Runs Bat	0.0734	0.0271	0.3692	0.6308
Strike Outs Bat	0.0840	0.0279	0.3321	0.6679
Walks Bat	0.1130	0.0674	0.5965	0.4035
Hits Pitch	0.0431	0.0230	0.5336	0.4664
On Base Pitch	0.0696	0.0459	0.6595	0.3405
Home Runs Pitch	0.0331	0.0164	0.4955	0.5045
Strike Outs Pitch	0.1150	0.0756	0.6574	0.3426
Walks Pitch	0.0436	0.0062	0.1411	0.8589
Average Bat			0.3811	0.6189
Average Pitch			0.4974	0.5026
Overall Average			0.4392	0.5608

Table A.20 – Learning vs. Streakiness. Including recent history in long-run ability. STCcoeff is the coefficient on the short-run history in our base regression. STCcoeffOL is the coefficient on the short-run history when the long-run ability variable is measured up to the current observation. Streak is the ratio of STCcoeffOL to STCcoeff, the fraction of the coefficient on the short-run history that remains after including the short-run history in the long-run variable. Learn is equal to $1 - \text{Streak}$; i.e., the fraction of the coefficient on the short-run history that is attributable to learning.

Statistic	STCcoeff	STCcoeff5	STCcoeff10	STCcoeff25	ST25ST
Hits Bat	0.0326	0.0345	0.0377	0.0331	1.0153
On Base Bat	0.0628	0.0508	0.0491	0.0416	0.6624
Home Runs Bat	0.0734	0.0675	0.0604	0.0559	0.7616
Strike Outs Bat	0.0840	0.0801	0.0843	0.0762	0.9071
Walks Bat	0.1130	0.0999	.0922	0.0745	0.6593
Hits Pitch	0.0431	0.0279	0.0272	0.0297	0.6891
On Base Pitch	0.0696	0.0465	0.0394	0.0354	0.5086
Home Runs Pitch	0.0331	0.0287	0.0245	0.0222	0.6707
Strike Outs Pitch	0.1150	0.0988	0.0788	0.0607	0.5278
Walks Pitch	0.0436	0.0508	0.0476	0.0424	0.9725
Average Bat					0.8012
Average Pitch					0.6737
Overall Average					0.7374

Table A.21 – Learning vs. Streakiness. STCcoeff is the coefficient on the short-run history in our base regression. STCcoeff5, STCcoeff10, and STCcoeff25 are the coefficient on the short-run history when this history is lagged by 5, 10, and 25 observations, respectively, from the current observation. ST25ST gives the ratio of STCcoeff25 to STCcoeff.



A.4 Simulations

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

Table A.22 – 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 (each with 2 million observations) from the specification given in (1) with R_{it}^L as the measure for the players current state and the windowed ability given by (3), the mean t-statistic is reported in parentheses.

Magnitude of Streakiness

Frequency	$\Delta = 0.01$	$\Delta = 0.02$	$\Delta = 0.04$	$\Delta = 0.05$	$\Delta = 0.10$
$\kappa = 1\%$	0.0007 (0.21)	0.0007 (0.20)	0.0037 (1.04)	0.0037 (1.03)	0.0184*** (5.22)
$\kappa = 5\%$	0.0017 (0.47)	0.0031 (0.88)	0.0138*** (3.91)	0.0217*** (6.18)	0.0821*** (24.16)
$\kappa = 10\%$	0.0018 (0.52)	0.0073** (2.07)	0.0280*** (7.99)	0.0435*** (12.51)	0.1509*** (46.36)

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

Table A.23 – Analysis of simulated data for players of homogeneous abilities: $\mu = 0.270$ and $T_{1/2} = 25$. The first number in each box is the mean estimated γ coefficient across $N = 100$ simulations (each with 2 million observations) from the specification given in (1) with R_{it}^L as the measure for the players current state and the windowed ability given by (3), the mean t-statistic is reported in parentheses.

Magnitude of Streakiness

Observations	$\Delta = 0.01$	$\Delta = 0.02$	$\Delta = 0.04$	$\Delta = 0.05$	$\Delta = 0.10$
$n = 10^3$	-0.0707 (-0.34)	-0.0475 (-0.22)	0.0132 (0.18)	-0.0084 (0.02)	0.0224 (0.23)
$n = 10^4$	-0.0065 (-0.10)	0.0024 (0.07)	0.0163 (0.35)	0.0278 (0.58)	0.0890* (1.90)
$n = 10^5$	-0.0000 (0.00)	0.0029 (0.19)	0.0155 (1.00)	0.0234 (1.50)	0.0819*** (5.44)

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

Table A.24 – This table illustrates the number of observations required to identify streakiness as it depends on the magnitude of streakiness (for fixed $\kappa = 0.05$). The first number in each box is the mean estimated γ coefficient across $N = 100$ simulations from the specification given in (1) with R_{it}^L as the measure for the players current state. The mean t-statistic is reported in parentheses. The ability is fixed over for each simulation at $\mu = 0.270$ and hence the ability control is absorbed by the constant and therefore omitted.

Magnitude of Streakiness

Observations	$\Delta = 0.01$	$\Delta = 0.02$	$\Delta = 0.04$	$\Delta = 0.05$	$\Delta = 0.10$
$n = 10^3$	0.01 (0.03)	0.00 (0.03)	0.03 (0.10)	0.01 (0.05)	0.04 (0.09)
$n = 10^4$	0.01 (0.05)	0.01 (0.05)	0.00 (0.08)	0.04 (0.11)	0.37 (0.61)
$n = 10^5$	0.00 (0.04)	0.04 (0.05)	0.10 (0.27)	0.22 (0.48)	0.99 (0.99)

Table A.25 – This table shows the fraction of the tests for which the null hypothesis (of $\gamma = 0$) is rejected as it depends on the magnitude of streakiness and the number of observations (for fixed $\kappa = 0.05$). The first number in each box is the fraction of the simulations that are rejected at a 1 percent confidence level. The number below it in parentheses is the fraction of the simulations rejected at a 5 percent level. The estimates for γ are obtained from the specification in (1) with R_{it}^L as the measure for the players current state. The ability is fixed for each simulation at $\mu = 0.270$ and hence the ability control is absorbed by the constant and therefore omitted.



B Supplemental Appendix (for online publication only)

	(1)	(2)	(3)	(4)	(5)	(6)
	Hit	Hr	Strikeout	OnBase	Walk	WalkEx
state	0.00522 (0.95)	0.0439*** (8.10)	0.0622*** (10.82)	0.0408*** (8.50)	0.0909*** (16.22)	0.0866*** (15.70)
pitcher_ability	0.525*** (31.18)	0.393*** (19.01)	0.797*** (51.60)	0.587*** (43.41)	0.796*** (53.23)	0.795*** (53.26)
samehand	-0.0182*** (-13.18)	-0.00777*** (-11.77)	0.0243*** (16.02)	-0.0333*** (-20.98)	-0.0233*** (-18.76)	-0.0234*** (-18.83)
batter_home	0.00939*** (9.99)	0.00208*** (4.61)	-0.00825*** (-9.74)	0.0134*** (14.66)	0.00694*** (12.52)	0.00692*** (12.41)
batter_state (hits)						0.00792* (2.32)
batter_state (hrs)						0.0798*** (9.47)
Observations	1045160	1045160	1045160	1250826	1250826	1243403

t statistics in parentheses

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

Table B.1 – Robustness: Measuring Ability using Batter Fixed Effects. This table gives estimates for the batter regressions where fixed effects are used to control for batter ability. The dependent variable is listed directly under the column number. Each column gives the estimates from the OLS regression given in equation (6) and using the baseline measure of the batter’s state as estimated by equation (3). Column (6) includes the batter’s state in hit in home runs to capture the endogenous response of the defense. Batters with fewer than 2,500 attempts in the dataset were excluded from the regressions. Standard errors are clustered at the batter level. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	Prop10	Prop2p5	Add10	Add2p5	Dist10	Dist2p5
hot	0.0128*** (8.74)	0.0186*** (6.93)	0.0132*** (9.34)	0.0181*** (6.43)	0.00870*** (6.04)	0.00753*** (3.51)
cold	-0.0111*** (-8.59)	-0.0147*** (-5.81)	-0.0112*** (-8.31)	-0.0158*** (-6.26)	-0.00990*** (-6.61)	-0.0143*** (-5.08)
Observations	1489346	1489346	1489346	1489346	1399046	1399046

t statistics in parentheses

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

Table B.2 – Robustness to thresholds. Dependent Variable: On Base (Batters). This table reports the results for the same regressions as given in Columns (4)-(6) (additive, proportional and distribution based thresholds for hot and cold) except using 10% and 2.5% thresholds as an alternative to the 5% thresholds used in Table A.7. The same control variable were included in the regressions but the coefficients are omitted from the above table. All regressions include stadium \times year dummy control variables.

	(1)	(2)	(3)	(4)	(5)	(6)
	Prop10	Prop2p5	Add10	Add2p5	Dist10	Dist2p5
hot	0.00436** (3.03)	0.0119*** (4.71)	0.00543*** (3.86)	0.0131*** (4.83)	0.00221 (1.45)	0.00522** (2.64)
cold	-0.00476** (-3.19)	-0.00719* (-2.43)	-0.00437** (-2.67)	-0.00858** (-2.90)	-0.00489** (-2.81)	-0.00634* (-2.49)
Stadium Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1192266	1192266	1192266	1192266	1113087	1113087

t statistics in parentheses

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

Table B.3 – Robustness to thresholds. Dependent Variable: Hits (Batters). This table reports the results for the same regressions as given in Columns (4)-(6) (additive, proportional and distribution based thresholds for hot and cold) except using 10% and 2.5% thresholds as an alternative to the 5% thresholds used in Table A.5. The same control variable were included in the regressions but the coefficients are omitted from the above table. All regressions include stadium \times year dummy control variables.

	(1) Hit	(2) Hr	(3) Strikeout	(4) OnBase	(5) Walk
state	0.0525*** (8.15)	0.0991*** (15.96)	0.127*** (19.74)	0.0876*** (15.01)	0.150*** (25.90)
batter_ability	0.363*** (16.06)	0.682*** (48.93)	0.759*** (74.40)	0.536*** (22.27)	0.654*** (39.30)
pitcher_ability	0.533*** (33.53)	0.377*** (20.34)	0.822*** (64.30)	0.602*** (44.61)	0.792*** (59.15)
samehand	-0.0135*** (-12.47)	-0.00454*** (-8.95)	0.0177*** (17.67)	-0.0246*** (-19.75)	-0.0161*** (-19.05)
batter_home	0.00914*** (10.72)	0.00207*** (5.27)	-0.00834*** (-11.20)	0.0136*** (16.44)	0.00680*** (13.06)
Observations	1154102	1154102	1154102	1400748	1400748

t statistics in parentheses

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

Table B.4 – Longer history lengths (Batters). This table reports the results for the same regressions as given in Columns (1) of Table A.5 across all 5 statistics except using alternative history lengths of $L=40$ to measure the batter's current state as given by equation (3). All regressions include stadium \times year dummy control variables.

	(1) Hit	(2) Hr	(3) Strikeout	(4) OnBase	(5) Walk
state	0.0104** (3.17)	0.0324*** (9.76)	0.0347*** (11.15)	0.0347*** (11.82)	0.0567*** (19.07)
batter_ability	0.376*** (15.98)	0.733*** (52.82)	0.838*** (89.40)	0.564*** (21.80)	0.727*** (38.85)
pitcher_ability	0.533*** (33.53)	0.377*** (20.34)	0.822*** (64.35)	0.601*** (44.54)	0.791*** (58.95)
samehand	-0.0134*** (-12.28)	-0.00448*** (-8.69)	0.0177*** (17.54)	-0.0247*** (-19.56)	-0.0163*** (-18.95)
batter_home	0.00905*** (10.69)	0.00201*** (5.23)	-0.00815*** (-11.20)	0.0133*** (16.48)	0.00660*** (13.13)
Observations	1154102	1154102	1154102	1400748	1400748

t statistics in parentheses

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

Table B.5 – Shorter history lengths (Batters). This table reports the results for the same regressions as given in Columns (1) of Table A.5 across all five statistics except using alternative history lengths of $L=10$ to measure the batter's current state as given by equation (3). All regressions include stadium \times year dummy control variables.

	(1) Hit	(2) Hr	(3) Strikeout	(4) OnBase	(5) Walk
state	0.0612*** (9.79)	0.0460*** (7.44)	0.139*** (21.68)	0.0883*** (15.04)	0.0661*** (10.62)
pitcher_ability	0.380*** (16.23)	0.267*** (11.52)	0.704*** (56.66)	0.429*** (19.43)	0.656*** (43.68)
batter_ability	0.562*** (49.15)	0.717*** (66.16)	0.831*** (80.94)	0.623*** (62.43)	0.629*** (66.39)
samehand	-0.0132*** (-10.44)	-0.00401*** (-8.62)	0.0182*** (10.05)	-0.0222*** (-17.45)	-0.0137*** (-18.28)
pitcher_home	-0.00865*** (-9.60)	-0.00151*** (-4.33)	0.00842*** (9.88)	-0.0117*** (-13.62)	-0.00533*** (-10.35)
Observations	1095673	1095673	1095673	1297611	1297611

t statistics in parentheses

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

Table B.6 – Longer history lengths (Pitchers). This table reports the results for the same regressions as given in Columns (1) of Table A.6 across all five statistics except using alternative history lengths of $L=40$ to measure the pitchers's current state as given by equation (3). All regressions include stadium \times year dummy control variables.

	(1) Hit	(2) Hr	(3) Strikeout	(4) OnBase	(5) Walk
state	0.0289*** (9.68)	0.0182*** (5.76)	0.0639*** (18.08)	0.0459*** (16.67)	0.0148*** (4.79)
pitcher_ability	0.393*** (16.27)	0.274*** (11.62)	0.767*** (60.96)	0.447*** (19.38)	0.691*** (45.14)
batter_ability	0.562*** (49.16)	0.716*** (66.20)	0.830*** (81.03)	0.622*** (62.42)	0.629*** (66.39)
samehand	-0.0131*** (-10.39)	-0.00401*** (-8.59)	0.0182*** (10.03)	-0.0222*** (-17.40)	-0.0137*** (-18.23)
pitcher_home	-0.00857*** (-9.61)	-0.00150*** (-4.31)	0.00829*** (9.98)	-0.0115*** (-13.62)	-0.00533*** (-10.32)
Observations	1095673	1095673	1095673	1297611	1297611

t statistics in parentheses

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

Table B.7 – Shorter history lengths (Pitchers). This table reports the results for the same regressions as given in Columns (1) of Table A.6 across all five statistics except using alternative history lengths of $L=10$ to measure the pitchers's current state as given by equation (3). All regressions include stadium \times year dummy control variables.