Testing the Minimax Theorem in the Field:

The Interaction between Pitcher and Batter in Baseball

Christopher Rowe

Advisor: Professor William Rogerson

Abstract

John von Neumann’s Minimax Theorem is a central result in game theory, but its practical applicability is questionable. While laboratory studies have often rejected its conclusions, recent field studies have achieved more favorable results. This thesis adds to the growing body of field studies by turning to the game of baseball. Two models are presented and developed, one based on pitch location and the other based on pitch type. Hypotheses are formed from assumptions on each model and then tested with data from Major League Baseball, yielding evidence in favor of the Minimax Theorem.

May 2013

MMSS Senior Thesis

Northwestern University
# Table of Contents

Acknowledgements .................................................. 3
Introduction .......................................................... 4
  The Minimax Theorem .............................................. 4
  Central Question and Structure .................................. 6
Literature Review .................................................... 6
  Laboratory Experiments .......................................... 7
  Field Experiments ................................................ 8
  Summary .......................................................... 10
Models and Assumptions ........................................... 10
  The Game .......................................................... 10
  Pitch Location Model ............................................ 13
  Pitch Type Model ................................................ 21
Hypotheses ........................................................... 24
  Pitch Location Model ............................................ 24
  Pitch Type Model ................................................ 31
Data Analysis ........................................................ 33
  Data .............................................................. 33
  Pitch Location Model ............................................ 34
  Pitch Type Model ................................................ 37
Conclusion ............................................................. 41
  Summary of Results .............................................. 41
  Future Research .................................................. 43
References ............................................................. 44
Appendix A ............................................................. 47
Appendix B ............................................................. 59
Acknowledgements

I would like to thank everyone who had a role in this paper’s completion. This begins with the Office of Undergraduate Research, who provided me with the funds necessary to complete this project, and everyone at Baseball Info Solutions, in particular Ben Jedlovec and Jeff Spoljaric, who provided me with data. I would like to especially thank Professor Jeffrey Ely for inspiring my interest in this particular topic, Kyra Sutton for providing edits, Derek Song for his help with Stata and econometrics, and Professor Joseph Ferrie for guiding me through the process. Finally, I would like to thank the MMSS program, especially Sarah Muir Ferrer and my advisor, Professor William Rogerson. To Sarah, thank you for always making the MMSS lounge a welcoming place. To Professor Rogerson, thank you for being my advisor, teacher, and program director for the past three years. I have your teaching to thank for my enthusiasm for game theory and your guidance to thank for the successful completion of this thesis.
INTRODUCTION

THE MINIMAX THEOREM

In any two-player zero-sum game of perfect information with finitely many strategies, there is an optimal mixed strategy solution. Furthermore, by performing according to this solution, each player maximizes her minimum possible payoff, and this payoff is unique in the sense that no other outcome involving optimal play can make either player better or worse off. This is what John von Neumann's Minimax Theorem tells us. The Minimax Theorem is an essential result in game theory. Many of the advances made in game theory over the last century have stemmed from von Neumann's proof in 1928. As von Neumann himself said, “there could be no theory of games [...] without that theorem” (Casti, 1996). In addition to its role within game theory, the Minimax Theorem has had an impact on the social sciences. Situations involving direct competition are relevant in economics, political science, and psychology. The Minimax Theorem provides us with the tools to analyze these situations.

There are reasons to trust the Minimax Theorem. We are familiar with the importance of unpredictability from our own experiences in competitive situations. For example, Rock-Paper-Scissors is a commonly played game where most people feel the need to vary their actions over time to mimic randomness. Deception is often a natural aspect of competition and playing a mixed strategy seems like a reasonable way to keep an opponent off-balance. Additionally, the suggestion that

\footnote{For more on the impact of the Minimax Theorem, see Casti, John L., \textit{Five Golden Rules}, especially pp. 16-19, 23-26.}
people maximize their minimum payoff sounds appealing. Most people are risk averse, so it makes sense that they would aim to minimize their potential loss. In many common games, such as Rock-Paper-Scissors, the equilibrium suggested by the Minimax Theorem agrees with our intuition.

However, there are a few reasons we might doubt the practical applicability of the Minimax Theorem. While appealing, mixed strategies may be implausible. In order for a mixed strategy solution to be played, each player must choose a strategy to make the other player indifferent. It is counterintuitive that players choose their strategies based only on their opponents’ payoffs. Furthermore, the performance of a mixed strategy requires genuine random selection of actions. Since we are not random number generators, this is a serious concern. Our attempts at randomization may be influenced by biases. Unsuccessful randomization leads to predictability, which can be exploited in competitive situations. There are suggestions that players engage in “k-level thinking” and try to stay one step ahead of their opponents to gain an advantage. This idea is supported by results in games such as the “two-thirds of the average game.”

Lastly, it is important to note that the equilibrium prescribed by the Minimax Theorem is highly unstable. If one player deviates slightly from his equilibrium strategy, the other has an incentive to change his strategy drastically, possibly even ceasing to play a mixture at all.

---

2 In the “two-thirds of the average game,” all players must choose an integer between 0 and 100, inclusive. The winner is the player who chooses the number closest to two-thirds of the average of all the integers. The Nash equilibrium solution is for every player to choose 0, but in practice the game does not play out this way. This result indicates that people may try to think a certain number of steps ahead rather than play the perfectly rational solution.
CENTRAL QUESTION AND STRUCTURE

In conclusion, it is unclear whether the Minimax Theorem can adequately describe human behavior. This is the central question that will be addressed in this paper. This study uses the interaction between the pitcher and the batter in the game of baseball to test whether experts perform as prescribed by the Minimax Theorem. The remainder of the paper will be structured as follows: we will first present the existing literature on the viability of the Minimax Theorem, differentiating between laboratory and field experiments. We will next present two separate models based on pitch location and pitch type along with justified assumptions about the strategies and payoffs. We will then use the Minimax Theorem to formulate hypotheses for each of the two models. Finally, we will test our hypotheses using empirical data and conclude.

LITERATURE REVIEW

The existing research on the viability of the Minimax Theorem can be divided into two separate categories: laboratory experiments and field experiments. Although initially only laboratory experiments were used to test the Minimax Theorem, field experiments have become popular over the past decade. Each type of study has its strengths and weaknesses. While laboratory experiments have clear results, they often suffer from a lack of external validity. On the other hand, field experiments provide greater external validity, but are often too complicated to draw clear conclusions. We need to keep this in mind while we review the research that has been conducted to date.
LABORATORY EXPERIMENTS

Laboratory experiments ease the process of developing models and analyzing data, but may lack external validity. The laboratory studies focusing on the Minimax Theorem follow the same general structure. The participants are typically presented with a simple but unfamiliar game with few strategies and clear payoffs. The game is designed so that the participants' actions are easily observable. Pairs of participants typically play the game repeatedly, generating a long sequence of trials. Thus the researchers are able to obtain an abundance of data influenced by a limited number of factors. This makes the process of analysis straightforward.

However, laboratory studies suffer from a possible lack of external validity. While the forced interactions that take place in these experiments are easy to analyze, they may not accurately reflect human behavior outside the laboratory. Here our interpretation of mixed strategies comes into play. Game theory is often used to model competitive situations in which the agents involved have years of experience to develop their abilities, so long-term behavior is of particular interest. Laboratory studies may tell us about how people behave immediately after introduction to a game, but the limited duration of these studies prevents conclusions about long-term behavior. Other issues may arise in laboratory studies as well. For instance, the incentives offered in laboratory experiments may be insufficient to ensure the payoffs assumed by researchers.

The laboratory experiments surrounding the Minimax Theorem have typically rejected it. The most well-known of these experiments, performed by O'Neill (1987), initially concluded in favor of the Minimax Theorem. However,
further examination of O’Neill’s own study by Brown and Rosenthal (1990) revealed that the study's results actually contradicted the Minimax Theorem. Additional studies (Rapoport and Boebel, 1992; Ochs, 1995) also provided mostly negative evidence. Furthermore, research on humans and randomness revealed a human tendency to fallaciously expect alternation in random sequences (Bar-Hillel and Wagenaar, 1991), giving weight to some of the concerns surrounding the Minimax Theorem. These laboratory experiments gave rise to alternative methods for modeling human behavior in games. Erev and Roth (1998) developed a model of adaptive play, in which players learn about their opponents over time and respond accordingly. Through empirical testing, they found that the predictions of their model outperformed the predictions of the Minimax Theorem.

FIELD EXPERIMENTS

In contrast to laboratory experiments, field experiments often have good external validity, but suffer from a lack of control. The field experiments on the Minimax Theorem have focused on professionals at work. Since the professionals observed in these studies are engaged in real-life human behavior, we can be reasonably certain that the conclusions of field experiments are generalizable. Unfortunately, we do not have the degree of control over field experiments that we have over laboratory experiments. The competitive situations studied in the field are often multi-faceted, which complicates the interpretation of any results that are obtained. Additionally, some strategies or preferences may be unobservable, which makes modeling and testing more challenging. Finally, there is often a lack of control over the quantity of data in the case of field experiments. These issues
sometimes make it difficult to arrive at reliable conclusions. The literature on the Minimax Theorem has recently moved towards field experiments because of their external validity, but they often lead to uncertain results.

Field experiments have generally offered positive evidence for the Minimax Theorem. In 2001, Walker and Wooders performed the first of these experiments by turning to the world of sports. They used the interaction between server and returner in tennis and found evidence mostly in favor of the Minimax Theorem, although they noted that players alternated too often between their actions.

Following their lead, later researchers used penalty kicks in soccer (Chiappori, Levitt, and Groseclose, 2002; Palacios-Huerta, 2003) and found positive results. A second study using tennis serves affirmed the results from Walker and Wooders and found that servers exhibited serial independence across serves (Hsu, Huang, and Tang, 2007). In an attempt to discover the reason behind the discrepancy between lab and field studies, Palacios-Huerta and Volij (2008) hypothesized that professionals have greater abilities in all (even unfamiliar) competitive situations due to their job experience. However, while Palacios-Huerta and Volij found evidence in favor of their hypothesis in the lab, replications of their study (Levitt, List, and Reiley, 2009; Wooders, 2010) have led to conflicting evidence.

Furthermore, a recent study by Kovash and Levitt (2009) has cast doubt on these earlier field studies. Pointing out the small data samples used in previous field studies, Kovash and Levitt turned to baseball and football. Using the choice between different pitch types in baseball and the choice between running and passing plays in football, they found evidence against the Minimax Theorem using much larger
data sets than those used in previous field studies. However, Kovash and Levitt did not present a fully developed model to support their results. Thus baseball is worth a re-examination.

**SUMMARY**

While the laboratory studies tend to reject the Minimax Theorem, the field studies tend to accept it. We might then conclude that people initially do not perform according to the Minimax Theorem, but develop the ability to do so in their professions over time. However, in addition to conflicting research on both sides, the relatively small body of research on the Minimax Theorem makes it difficult to reach a reliable conclusion at this time. This paper adds to the growing body of field experiments on the Minimax Theorem and in particular builds off the use of baseball in Kovash and Levitt (2009).

**MODELS AND ASSUMPTIONS**

**THE GAME**

Every event in baseball begins with an interaction between a single pitcher and a single batter in direct competition with each other. Baseball is by definition a team sport, but it has a highly individual focus. Although the pitcher and batter may be slightly influenced by the actions of fielders or baserunners, for the most part they interact as individuals independent of the other players in the game. This focus on individual competition is what allows us to model baseball as a two-player game.

We must first gain an understanding of the players in our game and how they interact. On every pitch, the pitcher and batter stand about 60 feet apart. Most generally, each player’s goal is to help his team win. The pitcher improves his
team’s chances of winning by preventing runs, while the batter improves his team’s chance of winning by producing runs. The pitcher and the batter then have directly conflicting goals. The pitcher varies the location, velocity, and movement of the ball to make it as difficult as possible for the batter to generate runs for his team. While different pitchers throw at different velocities, the fastest pitches reach home plate in less than one-half of one second, while even the slowest pitches reach home plate in well less than one second. Furthermore, the batter must make his decision before the ball reaches home plate in order to have any chance of making contact. Thus the batter has only a brief opportunity to recognize the trajectory of the ball and react.

What strategies can we extract from this situation? At the moment prior to the pitch, the pitcher aims for a particular location and chooses to throw the ball with some velocity and movement. The choice of the velocity and movement can be considered as a choice of pitch type: a fastball typically has greater velocity but less movement, while an off-speed pitch typically has greater movement but lower velocity. Except in extreme circumstances, the pitcher has complete control over pitch type, so in this case we are able to observe his choice. Pitchers also have some degree of control over pitch location. However, they are not perfectly accurate, so the choice of location is only partially observable. Meanwhile, the batter’s strategy is less clear. Once the pitch is thrown, he chooses whether or not to swing, but to some extent this is a reaction to the pitcher’s strategy. Since we want to model this situation as a simultaneous-move game, we want to consider the batter’s strategy prior to the pitch. We can describe the batter’s strategy based on his mental approach at this moment. Professional baseball players have affirmed that
anticipation is an important part of hitting. The batter’s strategy can then be described as the decision to anticipate a particular pitch and to prepare for that pitch in a certain way. Of course, this strategy is not perfectly observable. However, we may be able to partially observe the batter’s choice of strategy by observing whether or not he chooses to swing. Thus we have a thorough understanding of the strategies for each player. The pitcher’s strategy contains two basic elements: pitch type, which is observable, and pitch location, which is only partially observable. The batter’s strategy contains a single complex element: the choice of mental approach before the pitch, which may be partially observable. They may be slightly more complicated under certain circumstances, but these are essentially the strategies of the pitcher and the batter.

Before we move on to the process of modeling, we need to make a few observations about the payoffs. As previously discussed, the pitcher and batter are each attempting to maximize their own team’s chances of victory. Therefore, each player’s payoff for any particular event can be viewed as his team’s resultant probability of winning. Since the two teams’ probabilities of winning always sum to one, these payoffs make the interaction between pitcher and batter a zero-sum game. It is important to note that this interpretation somewhat complicates the

---

3 Ted Williams, who is in the Major League Baseball Hall of Fame, is famous for his dedication to the mental aspects of hitting. In his book *The Science of Hitting*, he wrote, “you’ve got to guess [...] how smart a guy is at the plate [...] that’s 50 per cent of it. From the ideas come the ‘proper thinking’ and the ‘anticipation,’ the ‘guessing.’”

4 This is similar to the characterization of the returner’s strategy presented by Walker and Wooders (2001). They described this strategy as a decision to anticipate a serve in one direction or another, which they called an “overplay” either to the left or to the right.
matter. Even between the same pitcher and batter, not all pitches will give rise to
the same payoff matrix. A number of factors, such as runners on base, balls, strikes,
outs, score, and inning all change how much a particular event influences each
team’s chances of winning. This will clearly affect modeling and testing, so it must
be taken into account.

We have now come to understand how we might generally model the
interaction between pitcher and batter as a game. While the complicated nature of
baseball creates some problems, we will take advantage of its complexity by
modeling the interaction between pitcher and batter in two different ways: first, we
will first present a model focused on pitch location and then we will present one
focused on pitch type. In both cases, we will explain the strategies for each player
and justify assumptions about the payoffs.

PITCH LOCATION MODEL

The first model we will explore is based on pitch location. The pitcher and
the batter are the only two players. We will first present the strategies available to
each player and then make assumptions about the payoff matrix. We will later use
these assumptions to form testable hypotheses for this model.

In this model, each player has two strategies. The pitcher can choose either
to attempt to throw the ball inside the strike zone or to attempt to throw the ball
outside the strike zone. The batter can choose either to take an “aggressive”
approach before the pitch or to take a “conservative” approach before the pitch. The
pitcher may often attempt to throw a ball outside the strike zone with the hope of
getting a strike. Nevertheless, we will call the pitcher’s strategy “strike” if he
attempts to throw the ball inside the strike zone and “ball” if he attempts to throw the ball outside the strike zone. Before we move on, we can offer a more detailed explanation of the batter’s strategy. As mentioned above, while the batter’s mental approach is important, to some extent the batter still has the ability to react to the pitch once it is thrown. We can then view the batter’s initial glimpse of the pitch as a signal. Each possible signal induces a probability distribution over all pitch locations. For example, the batter might get a signal that suggests the ball is most likely to end up in the low and inside part of strike zone. There will be some probability he has misread the pitch and it is actually outside the strike zone, or perhaps inside the strike zone but high, and so on for every possible pitch location.

We can then consider the batter’s strategy as a map, which, for every possible signal, chooses either “swing” (in which case the batter swings) or “take” (in which case the batter does not swing). Strategies that fall under the category of “aggressive” are those in which the batter swings for more signals, while strategies that fall under the category of “conservative” are those in which the batter swings for fewer signals. For example, a particular signal might suggest to the batter that the pitch will be inside the strike zone with probability .5 and outside the strike zone with probability .5. In an “aggressive” strategy, the batter might swing for this signal, while he might not swing for this signal in a “conservative” strategy. In a mixed strategy equilibrium, the batter will randomize across these maps. To recap, the pitcher has a choice between “strike” and “ball” (or S and B, respectively), while the batter has a choice between “aggressive” and “conservative” (or A and C, respectively).
As previously discussed, each player’s strategy has more components than this model takes into account. For example, pitch type is not considered in either the pitcher’s strategy or the batter’s strategy. However, this fact should not cast doubt on our ability to draw conclusions from this model. To see this, consider the strategy “strike” for the pitcher. We can consider “strike” as the group of all strategies in which the pitcher tries to throw the ball inside the strike zone.\(^5\) While this grouping may make certain aspects of the game less clear, it will not affect our ability to test for optimal behavior. If a pitcher consistently achieves better results from playing “strike” than from playing “ball,” he is not behaving optimally (as he would be better off if he played “strike” more frequently). The pitcher may also need to adjust another aspect of his strategy, such as pitch type, in order to achieve optimal play. However, what is important is the fact that the need for any adjustment reflects inefficient play. If both players were behaving optimally, the Minimax Theorem states that all strategies played with positive probability should produce the same expected payoff, which should then continue to hold for grouped strategies. Therefore, this way of describing each player’s strategies still allows us to test the Minimax Theorem.

We will assume that the strategies as specified truly exist for both players and are played in equilibrium. Clearly the pitcher will frequently try to throw a strike, and often the pitcher will try to get the batter to “chase” a ball. For the batter it is less obvious that both strategies exist, but for our purpose we merely need to show that some variation in aggressiveness exists. Although batters often look to

\(^5\) Walker and Wooders (2001) used this same tactic in their study on tennis serves. They ignored spin and other factors, grouping serves based on their location.
swing, batters will sometimes take a pitch (in other words, decide not to swing regardless of pitch location or pitch type), indicating that there is some variation in the batter’s approach. Furthermore, if we return to the concept of anticipation, it seems that a batter would take a less aggressive approach if he anticipated a ball than if he anticipated a strike. We can then separate the batter’s strategies into those that are more aggressive and those that are more conservative. The exact division of these strategies is not necessary information as long as we know that some division is possible. The fact that we know both strategies are played indicates they are both played in equilibrium.

We will now consider the payoffs in our model. Each strategy profile induces a probability distribution over all possible outcomes of an at-bat. For instance, if the pitcher chooses to play “strike” and the hitter chooses to play “aggressive,” this generates a certain probability of a home run, a certain probability of a strike, and so on for all other possible outcomes. By taking into account the impact of each possible outcome on win probability, we can more simply represent this result as the pitching team’s expected probability of winning after the pitch.\(^6\) As previously discussed, since the pitcher and batter each have a considerable stake in their respective team’s success, it is clear that the batter will attempt to minimize this probability while the pitcher will attempt to maximize it. The matrix for the game can be seen below (Figure 1), where \(\pi_{xy}\) denotes the pitching team’s probability of

\(^6\) Since this is a zero-sum game, only one player’s payoff is necessary to specify the payoffs of both players.
winning when the pitcher plays X and the batter plays Y. We should recall that these payoffs differ depending on the particular pitcher, batter, and game situation.

With the basics of our model completed, we can now develop some natural assumptions about the payoffs. We will present and justify eight assumptions that will hold for every possible game situation and pair of players. Some of the assumptions will compare payoffs from situations that are identical with the exception of the number of balls. Specifically, these assumptions will compare 2-2 counts and 3-2 counts. When this is the case, $\pi_{XY}^1$ will denote the payoff that results from the pitcher playing X and the batter playing Y on a 2-2 count, while $\pi_{XY}^2$ will denote the payoff that results from the pitcher playing X and the batter playing Y on a 3-2 count. The eight assumptions are presented below in Figure 2.

<table>
<thead>
<tr>
<th>Assumption (1)</th>
<th>$\pi_{BA} &gt; \pi_{BC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption (2)</td>
<td>$\pi_{SC} &gt; \pi_{SA}$</td>
</tr>
<tr>
<td>Assumption (3)</td>
<td>$\pi_{SC} &gt; \pi_{BC}$</td>
</tr>
<tr>
<td>Assumption (4)</td>
<td>$\pi_{RA} &gt; \pi_{SA}$</td>
</tr>
<tr>
<td>Assumption (5)</td>
<td>$\pi_{BA}^2 - \pi_{BC}^2 &gt; \pi_{BA}^1 - \pi_{BC}^1$</td>
</tr>
<tr>
<td>Assumption (6)</td>
<td>$\pi_{SC}^1 - \pi_{SA}^1 &gt; \pi_{SC}^2 - \pi_{SA}^2$</td>
</tr>
<tr>
<td>Assumption (7)</td>
<td>$\pi_{SC}^2 - \pi_{BC}^2 &gt; \pi_{SC}^1 - \pi_{BC}^1$</td>
</tr>
<tr>
<td>Assumption (8)</td>
<td>$\pi_{BA}^1 - \pi_{SA}^1 &gt; \pi_{BA}^2 - \pi_{SA}^2$</td>
</tr>
</tbody>
</table>

---

7 The count is the number of balls and strikes. When a count is listed as 2-2 or 3-2, the number of balls comes first and the number of strikes comes second.
Assumptions (1) and (2) are straightforward. Both of these assumptions come from considering the perspective of the batter. Assumption (1) states that if the pitcher attempts to throw a ball, the batter will on average achieve a better result by being conservative. Assumption (2) states that if the pitcher attempts to throw a strike, the batter will on average achieve a better result by being aggressive. Although these assumptions are intuitive, we will provide justification for both. The strike zone is positioned so that the batter can more easily reach strikes than balls. Thus the batter is better off swinging at strikes than balls, and he is better off taking balls than strikes. Since the pitcher has some degree of control over his pitches, he is more likely to throw a ball when he plays “ball” and he is more likely to throw a strike when he plays “strike.” The fact that both strategies are played in equilibrium completes the justification of these two assumptions. If Assumption (1) were false, the batter would be better off playing “aggressive” when the pitcher plays “ball.” When the pitcher plays “strike,” the batter would see more strikes and fewer balls, which would make playing “aggressive” more appealing and “conservative” less so. Then “aggressive” would be a dominant strategy for the batter, which cannot be the case since “conservative” is played in equilibrium. Assumption (2) is true by similar logic. If Assumption (2) were false, “conservative” would be a dominant strategy for the batter, which cannot be the case since “aggressive” is played in equilibrium.

Assumptions (3) and (4) follow similar logic, but require the perspective of the pitcher rather than the batter. Assumption (3) states that the pitcher is better off playing “strike” if the batter is being conservative, while Assumption (4) states that the pitcher is better off playing “ball” if the batter is being aggressive. If
Assumption (3) were not true, the pitcher would be better off playing “ball” when the batter plays “conservative.” However, the batter is more likely to chase if he is playing “aggressive,” making “ball” an even more appealing strategy for the pitcher. Therefore, “ball” would be a dominant strategy for the pitcher, which we know is not true since both “strike” and “ball” are played in equilibrium. Finally, under the rest of our assumptions, if Assumption (4) did not hold, “strike” would be a dominant strategy for the pitcher. This is impossible, so Assumption (4) must hold.

Assumptions (5)-(8) can be justified by comparing the situation that arises as the count moves from 2-2 to 3-2. Balls are always bad for the pitcher.\(^8\) Therefore, as more balls are thrown, the pitcher’s probability of winning decreases for all possible strategy profiles. We can identify which strategy profiles will be affected the most from the increased number of balls by considering the particular nature of the distribution induced by each strategy profile. For any particular strategy profile, the move from a 2-2 count to a 3-2 count can affect payoffs only through the increased probability of a walk.\(^9\) Therefore, payoffs will be most influenced by the extra ball in 3-2 counts for whichever strategy profile induces the highest probability of a ball. This will clearly be the case when the pitcher plays “ball” and the batter plays “conservative.” If the batter plays “aggressive,” he is more likely to swing, thus making a ball less likely. If the pitcher attempts to throw a strike, a ball

\(^8\) Pitchers do occasionally intentionally walk batters. However, with the possible exception of these relatively uncommon events, balls should always decrease the pitching team’s chances of winning.

\(^9\) Although batters might fare differently on balls hit in play on 3-2 counts, this is the result of each player’s choice of action. Here we are considering the change in payoffs for a particular strategy profile, so the players’ choices are irrelevant.
will clearly be less likely. Using this reasoning, we can also conclude that a ball will be least likely when the pitcher plays “strike” and the batter plays “aggressive.” Therefore, in 3-2 counts, $\pi_{BC}$, $\pi_{BA}$, $\pi_{SC}$, and $\pi_{SA}$ will all decrease relative to 2-2 counts, but $\pi_{BC}$ will decrease the most, while $\pi_{SA}$ will decrease the least. A simple manipulation of these two facts gives us Assumptions (5)-(8) listed above.\(^\text{10}\)

These assumptions make intuitive sense as well. Consider first Assumptions (5) and (6). We can think of $\pi_{BA} - \pi_{BC}$ as the advantage the batter gains from being “right” on a ball and $\pi_{SC} - \pi_{SA}$ as the advantage the batter gains from being “right” on a strike. In this context, our assumptions are equivalent to the following statement: on 3-2 counts compared to 2-2 counts, the batter gains a greater advantage from being right when the pitcher plays “ball” (Assumption (5)), but a lesser advantage from being right when the pitcher plays “strike” (Assumption (6)). When batters are right on an attempted ball on 3-2 counts, they get the extra advantage of a likely walk as opposed to merely an extra pitch with a more favorable count. In the case of an attempted strike, guessing wrong still produces a small possibility that the pitcher misses his target. Thus by being right on an attempted strike, the batter loses out on that possibility of a ball. On 3-2 counts, balls are more rewarding, making it less worthwhile to guess right on an attempted strike.

Assumptions (7) and (8) are similarly intuitive. We can view $\pi_{SC} - \pi_{BC}$ as the advantage the pitcher gains by being right when the batter plays “conservative” and $\pi_{BA} - \pi_{SA}$ as the advantage the pitcher gains by being right when the batter plays

\(^{10}\) See Appendix A.1 for the explicit derivation of Assumptions (5)-(8).
“aggressive.” On 3-2 counts, the increased possibility of a walk makes it more important for the pitcher to be right when the batter plays “conservative.” This gives us the intuition behind Assumption (7). On the other hand, on 3-2 counts, it is less beneficial to guess right when the batter plays “aggressive” because guessing right in that situation involves attempting to throw a ball. If the batter manages to avoid swinging despite playing “aggressive,” the pitcher suffers more than he would on a 2-2 count. This is the intuition behind Assumption (8).

This concludes my justification of the eight assumptions on the pitch location model. These assumptions will be used in the next section to develop testable hypotheses.

**PITCH TYPE MODEL**

We will now consider a second model based on pitch type. As in the pitch location model, we will describe the strategies and payoffs. It is more difficult to form justifiable assumptions in this model. However, this model has advantages over the pitch location model that we will see when we form our hypotheses.

Once again, each player has two strategies. The pitcher can choose to throw either a fastball or an off-speed pitch, while the batter can choose to look for either a fastball or an off-speed pitch. We will refer to these strategies as “fastball” and “off-speed” for both players. The division between a fastball and an off-speed pitch was briefly mentioned earlier. A fastball is a fast pitch which typically takes a fairly straight path to the batter. An off-speed pitch is any pitch thrown at significantly reduced velocity, often with movement. While most fastballs are referred to simply as “fastballs,” cut fastballs are a more specific example. On the other hand, there are
many different types of off-speed pitches. Changeups, curveballs, and sliders are the most common. Every type of pitch can be labeled as either a fastball or off-speed pitch in this way. The batter’s strategy requires less interpretation in this case than it did in the pitch location model. We previously discussed the importance of anticipation for the batter. The strategy “fastball” for the batter then denotes all the strategies where the batter anticipates a fastball, while the strategy “off-speed pitch” for the batter denotes all the strategies where the batter anticipates an off-speed pitch. It is important to point out that we do not necessarily mean that the batter will take a guess at a certain pitch and react with certainty according to that guess. While this is a possibility in some circumstances, it is more likely that batters prepare for a particular pitch and adjust once the pitch is thrown. To recap, both the pitcher and the batter have a choice between “fastball” and “off-speed pitch” (or F and O, respectively), although these strategies mean slightly different things for each player.

As was the case in the pitch location model, this model’s grouping of strategies leads to some loss in information. While virtually all pitchers throw at least one type of fastball in addition to at least one type of off-speed pitch, different pitchers throw different types of fastballs and off-speed pitches. For example, one pitcher might throw a changeup as his only off-speed pitch, while another might throw a slider and a curveball. Furthermore, pitch location is unaccounted for in this model. However, as in the case of pitch location, this fact should not be an issue. Once again, our decision to group strategies into only two categories should not prevent the application of the Minimax Theorem. If a pitcher achieves worse results
from throwing off-speed pitches, then he is clearly not behaving optimally. As noted in the previous model, while the solution might require changes in other aspects of the pitcher’s strategy, what is important is that some adjustment is needed. Therefore, this division of strategies is a valid way to model baseball as a game for empirical testing.

We assume that the strategies specified in this model exist and are played in equilibrium. This is clear for the pitcher’s strategy since the differences between fastballs and off-speed pitches can be seen with the naked eye. Nearly every pitcher throws both fastballs and off-speed pitches. While it is less clear for the batter’s strategy, we can rely on the testimony of professional baseball players mentioned earlier in this paper. If their claims are truthful, batters rely on anticipating a particular pitch, and anticipating a pitch would certainly involve anticipating a certain pitch type. Thus it seems batters play both strategies specified above, indicating that they exist and are played in equilibrium.

Our interpretation of payoffs is the same as it was in the first model. Each strategy profile induces some probability distribution over outcomes, which can be summarized by the resultant expected win probability for the pitching team. The payoff matrix for this model is shown below in Figure 3. Once again, \( \pi_{xy} \) denotes the payoff that results when the pitcher plays \( X \) and the batter plays \( Y \). Finally, this matrix depends on the game situation and pair of players.

\[
\begin{array}{c|cc}
& F & O \\
\hline
F & \pi_{FF} & \pi_{FO} \\
O & \pi_{OF} & \pi_{OO} \\
\end{array}
\]

\( \text{Figure 3} \)
We could make a few basic assumptions about the payoff matrix, but none of them would help us form testable hypotheses. However, this model benefits from the fact that we can perfectly observe the pitcher’s choice of pitch type. As we will see in the next section, this will allow us to directly apply the Minimax Theorem to form hypotheses.

**HYPOTHESES**

Now that we have two fully developed models with all of the assumptions we need, we can move on to the process of forming hypotheses. We will begin with the pitch location model before moving on to the pitch type model. For both models, we will first present a characterization of equilibrium. We will then use the equilibrium solution together with our assumptions on each model in order to formulate testable hypotheses.

**PITCH LOCATION MODEL**

We will first consider the pitch location model. Since our game is a finite, two-player, zero-sum game with perfect information, there is a unique solution in mixed strategies. We can write the pitcher’s strategy as \((p_S, p_B)\) where \(p_S\) and \(p_B\) are the probabilities that the pitcher plays \(S\) and \(B\), respectively, and the batter’s strategy as \((h_A, h_C)\) where \(h_A\) and \(h_C\) are the probabilities that the batter plays \(A\) and \(C\), respectively. Since \(p_S + p_B = 1\) and \(h_A + h_C = 1\), we can fully describe each player’s equilibrium strategy with just one of these two probabilities. Through
some straightforward manipulation, we can arrive at the equilibrium solution
displayed in Figure 4.\textsuperscript{11}

\[
p_S = \frac{\pi_{BA} - \pi_{BC}}{\pi_{BA} - \pi_{BC} + \pi_{SC} - \pi_{SA}}
\]

\[
h_A = \frac{\pi_{SC} - \pi_{BC}}{\pi_{SC} - \pi_{BC} + \pi_{BA} - \pi_{SA}}
\]

Figure 4

We will now present four separate hypotheses. We will first conclude that
pitchers should throw inside the strike zone more on 3-2 counts than on 2-2 counts.
We will next develop a few additional assumptions and conclude that batters should
swing more on 3-2 counts than on 2-2 counts. We will then conclude that batters
should swing more on 3-2 counts than on 2-2 counts for a given pitch location.
Finally, we will conclude that the outcome where the pitcher throws a strike and the
batter swings (denoted by (Strike, Swing) from this point on) should be more
common in 3-2 counts than on 2-2 counts, while the outcome where the pitcher
throws a ball and the batter does not swing (denoted by (Ball, Take) from this point
on) should be less common. These hypotheses are summarized below in Figure 5.
For each hypothesis, we will provide a general justification\textsuperscript{12} and briefly discuss any
relevant issues.

| H1 | \textit{Pitchers throw inside the strike zone more often on 3-2 counts than on 2-2 counts} |
| H2 | \textit{Batters swing more often on 3-2 counts than on 2-2 counts} |
| H3 | \textit{Batters swing more on 3-2 counts than on 2-2 counts for a given pitch location} |
| H4 | \textit{The outcome (Strike, Swing) is more common on 3-2 counts than on 2-2 counts and the outcome (Ball, Take) is less common on 3-2 counts than on 2-2 counts} |

Figure 5

\textsuperscript{11} The computation behind the equilibrium solution can be found in Appendix A.2.
\textsuperscript{12} The formal justification behind each hypothesis can be found in Appendix A.
We will first consider the hypothesis that pitchers throw inside the strike zone more on 3-2 counts than on 2-2 counts (from this point forward, we will refer to this hypothesis as H1). We can manipulate our equilibrium solution using our assumptions on the payoffs to show that pitchers will play “strike” more often on 3-2 counts than on 2-2 counts.\(^{13}\) Since we assume pitchers have some degree of control over pitch location, we can use this claim to formulate H1.\(^{14}\)

The theory behind this hypothesis is not difficult to see. In 3-2 counts, the batter has an increased incentive to play “conservative.” If the pitcher uses his equilibrium strategy for 2-2 counts on 3-2 counts, the batter would achieve optimal play by always playing “conservative.” To return the batter to indifference, the pitcher must then play “strike” with greater frequency to incentivize “aggressive” relative to “conservative” for the batter.

This hypothesis is also fairly intuitive. It seems natural that the pitcher would be more wary of throwing a ball on a 3-2 count than on a 2-2 count. Pitchers might attempt to throw strikes more on 3-2 counts even if they do not behave optimally. While this hypothesis is still worth testing, it will not be as informative as some of our other hypotheses.

We will now consider the hypothesis that batters swing more on 3-2 counts than on 2-2 counts (from this point forward, we will refer to this hypothesis as H2). We can perform the same manipulation as in the case of H1 to conclude that batters

\(^{13}\) The formal justification for this claim is found in Appendix A.4.

\(^{14}\) In these situations, there can be concerns about aggregation. However, our claims hold for every pitcher, batter, and game situation, and should thus survive aggregation. See Chiappori, Levitt, and Groseclose (2002) for more on the issue of aggregation.
will play “aggressive” more on 3-2 counts than on 2-2 counts.\textsuperscript{15} However, we cannot immediately formulate H2 from this claim as we did in the case of H1 because we do not directly observe the batter’s strategy. In the case of the pitcher’s strategy, we could depend on the fact that the pitcher has some control over pitch location. In the case of the batter’s strategy, the pitcher’s choice becomes a complicating factor. The batter might play “aggressive,” but if the pitch is far from the strike zone, the batter is unlikely to swing. On the other hand, the batter might play “conservative,” but if the pitcher throws the ball in the middle of the strike zone, the batter might swing anyway. Thus we need to develop assumptions about the batter’s probability of swinging in order to justify H2.

We will now investigate the batter’s probability of swinging and make some basic assumptions about his probability of swinging for different strategy profiles. As previously stated, each strategy profile induces a probability distribution across all possible outcomes. We can group all of these outcomes into two categories: outcomes in which the batter swings and outcomes in which he does not. Each strategy profile then results in some probability that the batter swings. We will let $s_{XY}$ denote the probability that the batter swings when the pitcher plays X and the batter plays Y. Now, letting $q_{XY}$ denote the probability that the pitcher plays X and the batter plays Y, we can describe the batter’s total probability of swinging. For a given pitcher, batter, and game situation, $s_{SA}$, $s_{SC}$, $s_{BA}$, and $s_{BC}$ are fixed. Therefore, we can represent the batter’s total probability of swinging as a function of $q_{SA}$, $q_{SC}$, $q_{BA}$, and $q_{BC}$, which is the first function displayed below in Figure 6. Since we know

\begin{footnotesize}
\textsuperscript{15} The formal justification for this claim is found in Appendix A.5.
\end{footnotesize}
that \( q_{xy} = p_x h_y \), we can rewrite the first function in Figure 6 in terms of \( p_S \) and \( h_A \).

The rewritten version is the second function that appears in Figure 6.

\[
\begin{align*}
\frac{f(q_{SA}, q_{SC}, q_{BA}, q_{BC})}{f(p_S, h_A)} &= q_{SA}s_{SA} + q_{SC}s_{SC} + q_{BA}s_{BA} + q_{BC}s_{BC} \\
&= p_S h_A s_{SA} + p_S (1-h_A)s_{SC} + (1-p_S)h_A s_{BA} + (1-p_S)(1-h_A)s_{BC}
\end{align*}
\]

**Figure 6**

We will now outline some basic assumptions on the batter’s probability of swinging for a given strategy profile. According to our definition of the batter’s strategy, the batter is more likely to swing for a given strategy of the pitcher if he plays “aggressive.” We will also assume that the batter is more likely to swing when the pitcher plays “strike.” We previously pointed out that the strike zone is positioned so that the batter can reach strikes more easily than balls. The batter will then be more likely to swing on strikes, which are more likely when the pitcher plays “strike”. Therefore, we can conclude that \( s_{SA} \) will be the highest of the four swing probabilities, while \( s_{BC} \) will be the lowest.

We can now formally justify H2. We can manipulate \( f(p_S, h_A) \) to show that it is strictly increasing in both \( p_S \) and \( h_A \).\(^{16}\) However, we previously demonstrated that in equilibrium the pitcher plays “strike” more often while the batter plays “aggressive” more often. Therefore, the batter’s total probability of swinging is higher for 3-2 counts than it is for 2-2 counts. This allows us to conclude that batters swing more on 3-2 counts than on 2-2 counts, completing our justification of H2.

We will now consider the hypothesis that batters swing more on 3-2 counts than on 2-2 counts for any given pitch location. For this hypothesis, we need to

\(^{16}\) This manipulation appears in Appendix A.6.
consider our interpretation of the batter’s strategy. We previously described the batter’s strategy as a map from the set of all signals to either “swing” or “take.” Here we will further develop this interpretation. We introduced conservative strategies with the concept of taking a pitch regardless of location and type. In these situations, the batter chooses not to swing for any signal. The batter might instead take a slightly less conservative approach and swing for the one signal that gives him the best expected outcome. As we move towards more aggressive strategies, the batter will swing for more signals, but we would expect him to continue swinging for the most appealing signals. Under the assumption that this is true, we can formulate H3. We previously showed that the batter will play “aggressive” more frequently in 3-2 counts than in 2-2 counts. Our interpretation of the batter’s strategy reveals that, for any given pitch location, the batter’s probability of swinging is greater the more he plays an aggressive strategy. Therefore, for any given pitch location, the batter swings more on 3-2 counts than on 2-2 counts.

The theory behind H2 and H3 is straightforward. In a 3-2 count, the pitcher has less incentive to play “ball.” Thus if the batter did not adjust his strategy, the pitcher would play the pure strategy “strike.” The batter must then play “aggressive” more frequently to incentivize “ball” relative to “strike.” While the theory behind these hypotheses is not difficult, the result is somewhat counterintuitive. We might expect the batter to swing less on 3-2 counts, because if the pitch is in fact a ball, his reward is greater than it would be on a 2-2 count. This is especially counterintuitive in the case of H3 because we control for the actual location of the pitch. In the case of H2, batters might swing more on 3-2 counts even
without playing a more aggressive strategy because pitchers throw the ball in the
strike zone more often (as we concluded in H2). H3 states that batters will swing
more even given that the pitch is outside the strike zone, depriving themselves of a
walk in the process. While H2 and H3 might still hold in the presence of non-
optimal play, it is not likely because of their counterintuitive nature.

Finally, we will consider the hypothesis that on 3-2 counts compared to 2-2
counts, the outcome (Strike, Swing) is more common while the outcome (Ball, Take)
is less common (we will refer to this hypothesis as H4). This conclusion follows
from the rest of the conclusions in this section. On 3-2 counts compared to 2-2
counts, the batter will play “aggressive” more often. From our interpretation of the
batter’s strategy, this means he will swing at a higher frequency for all given pitch
locations, including strikes. Furthermore, the pitcher will throw strikes more on 3-2
counts than on 2-2 counts. Thus on 3-2 counts, the batter will swing more often at
strikes, which will be thrown more often, leading to a higher frequency of the
outcome (Strike, Swing). Similarly, on 3-2 counts, the batter will take less often for
all given pitch locations and the pitcher will throw balls less often. Thus the
outcome (Ball, Take) will occur with lower frequency. While this hypothesis does
not add much conceptually, it provides us with another opportunity to test for
optimal play.

We have derived four testable hypotheses from the pitch location model
using our justified assumptions. H2 and especially H3 are of particular interest
because they are somewhat counterintuitive results. While we will test all four
hypotheses in the next section, our tests of H2 and H3 will be particularly revealing.
**PITCH TYPE MODEL**

We will now consider the pitch type model. Since this game is a two-player, zero-sum game with finitely many strategies, we can characterize the unique equilibrium solution as we did for the pitch location model. We can fully describe the equilibrium strategies of each player with the probabilities $p_F$ and $h_F$, where $p_F$ is the probability that the pitcher throws a fastball and $h_F$ is the probability that the batter anticipates a fastball. A simple computation yields the equilibrium solutions displayed in Figure 7.17

\[
p_F = \frac{\pi_{OF} - \pi_{OO}}{\pi_{FO} - \pi_{FF} + \pi_{OF} - \pi_{OO}}
\]
\[
h_F = \frac{\pi_{FO} - \pi_{OO}}{\pi_{FO} - \pi_{OO} + \pi_{FO} - \pi_{FF}}
\]

Figure 7

We will now produce two separate hypotheses for a given batter and game situation. We will first conclude that pitchers should achieve equivalent results for fastballs and off-speed pitches, and we will then conclude that pitchers should exhibit serial independence in their pitch selection. These hypotheses are summarized below in Figure 8. We will justify both of these hypotheses with the Minimax Theorem.

| H5 | On average, pitchers achieve equivalent outcomes for fastballs and off-speed pitches for a given batter and game situation |
| H6 | Pitchers exhibit serial independence in their pitch selection |

Figure 8

17 The computation behind the equilibrium solution can be found in Appendix A.3, but it is nearly identical to the computation for the pitch location model.
We will first consider the hypothesis that pitchers achieve equivalent results for fastballs and off-speed pitches (which we will refer to as H5). The Minimax Theorem requires players to be indifferent between all of their strategies. Thus for each player, every action played with positive probability must guarantee the same expected payoff. We can apply this claim to the pitch type model. A pitcher who throws both fastballs and off-speed pitches in equilibrium must be indifferent between the two choices. If, for example, fastballs achieved consistently better results in a particular situation than off-speed pitches, an optimally-behaving pitcher would strictly prefer to throw fastballs in that situation. Therefore, in equilibrium, neither fastballs nor off-speed pitches should achieve consistently better results. We should be able to observe this empirically. However, the payoff matrix differs across different pitchers, batters, and game situations. We must control for factors that can affect the payoff matrix when we test H5 using data.\textsuperscript{18}

We will now consider the hypothesis that pitchers exhibit serial independence in their pitch selection (which we will refer to as H6). The Minimax Theorem requires true randomization. Thus if a zero-sum game is played repeatedly, neither player’s actions should be correlated over time. We can apply this claim to the pitch type model as well. Pitchers should practice true randomization if they are behaving optimally. If a particular pitcher threw an off-speed pitch every time after he threw a fastball, an optimally-behaving batter would take note of this pattern and exploit it to his advantage. Pitchers should then exhibit serial independence in their strategies to avoid predictability. Since we perfectly
\textsuperscript{18}Kovash and Levitt (2009) addressed the problem of heterogeneity by controlling for confounding factors. We will follow their lead here.
observe the pitcher’s choice of pitch type, H6 is a testable hypothesis. Once again, payoffs differ across different pitchers, batters, and game situations. We will address this problem in empirical testing by controlling for possible confounding factors.

**DATA ANALYSIS**

We will now consider our tests for the six hypotheses set out in the previous section. We reject our hypothesis that pitchers exhibit serial independence, but fail to reject the other five hypotheses. These results largely support the Minimax Theorem.

**DATA**

Our data set includes every pitch thrown in the 2012 Major League Baseball regular season. The data was purchased from Baseball Info Solutions,\(^{19}\) who are able to gather a wealth of information on every pitch thrown using advanced technology. For each pitch, the data set identifies the pitcher, batter, game situation (home and away team, inning, score, baserunners, outs, and count), pitch type, pitch location, pitch result, and the result of the at-bat in which the pitch was thrown. The data set includes 705,321 pitches, 660 pitchers, and 960 batters. The pitchers range from 4 to 3,768 pitches thrown and from 2 to 956 batters faced, while the batters range from 1 to 2,921 pitches seen and from 1 to 741 plate appearances.\(^{20}\) There are 72,879 different pitcher-batter pairs, who match up for as many as 94 pitches and 22 at-bats, while some pitcher-batter pairs match up for only a single pitch. The

\(^{19}\) Kovash and Levitt (2009) used data from the same company, but from a different time span.

\(^{20}\) A “plate appearance” is any turn at bat for a particular player.
data set identifies 9 different possible pitch locations: In the Strike Zone, Outside, High, Inside, Low, High Outside, Low Outside, High Inside, and Low Inside. Pitches inside the strike zone fall only into the first category, while pitches outside the strike zone can fall into any of the others. The data set identifies 8 different pitch types: Fastball, Cut Fastball, Slider, Curveball, Changeup, Split Finger, Knuckleball, and Screwball. Pitches labeled as Fastball or Cut Fastball are considered fastballs, while the remaining pitches are considered off-speed pitches. 4,210 pitches were labeled “No Video” and are therefore eliminated whenever pitch type is relevant. The data set presents 50 different pitch results, all of which specify whether or not the batter swung. Some basic summary statistics can be found in Table 1, Table 2, and Table 3 of Appendix B. We use this data to test the hypotheses laid out in the previous section.

**PITCH LOCATION MODEL**

We will begin with our tests for the hypotheses of the pitch location model. Testing these hypotheses requires restricting our data to only those pitches thrown on 2-2 or 3-2 counts. Some summary statistics for these pitches can be found in Table 4. Our restricted data set contains 87,604 pitches, 12.42 percent of our original data set. 55,158 pitches (62.96 percent of our restricted data set) come from 2-2 counts, while 32,446 pitches (37.04 percent of our restricted data set) come from 3-2 counts. Although testing H1-H4 requires using a reduced data set, we still have a large number of observations for both 2-2 and 3-2 counts.

The results from the tests of hypotheses H1-H4 are presented in Appendix B. Table 5 contains the results for our test of H1, Table 6 contains the results for our
tests of H2 and H3, and Table 7 contains the results for our test of H4. With a slight exception in the case of H3, we fail to reject our hypotheses in all situations.

Table 5 shows that 41.83 percent of 2-2 pitches and 52.96 percent of 3-2 pitches were thrown in the strike zone. This difference is significant at a 1-percent level. We then see that pitchers do in fact throw the ball in the strike zone with greater frequency on 3-2 counts than on 2-2 counts, so H1 is supported by the data. As previously noted, this result is not surprising. It may be the case that pitchers would throw the ball in the strike zone more on 3-2 pitches even if they were not behaving optimally. We will now move on to H2 and H3 for more interesting results.

Table 6 shows, for both 2-2 counts and 3-2 counts, the frequency that the batter swings for a variety of pitch locations. First, we see that batters swing 74.20 percent of the time on 3-2 counts compared to 66.11 percent of the time on 2-2 counts. This difference is significant at a 1-percent level. Thus batters do in fact swing more on 3-2 counts than on 2-2 counts, which supports H2. For pitches inside the strike zone, batters swing 92.42 percent of the time on 3-2 counts and 90.04 percent of the time on 2-2 counts. For pitches outside the strike zone, batters swing 53.70 percent of the time on 3-2 counts compared to 48.90 percent of the time on 2-2 counts. Both of these differences are significant at a 1-percent level, which supports H3 for these two general pitch locations. Because of the specificity offered by the data, we can break down pitches outside the strike zone into eight different

---

21 All tests for equal frequencies are conducted using a simple one-sided t-test.
categories. For “High” pitches, batters actually swing significantly more frequently on 2-2 counts, which contradicts H3. For “Low” pitches, batters also swing more frequently on 2-2 counts, but the difference is not significant. For the remaining six locations, the batter swings more frequently on 3-2 counts, offering support of H3. Four of these locations obtain significance at a 5-percent level, while two of these four obtain significance at a 1-percent level as well. Thus with only one significant exception, H3 is supported by empirical testing. In summary, the evidence for H2 is unqualifiedly positive, while the evidence for H3 is mostly positive. We therefore fail to reject either H2 or H3. As discussed in the hypotheses section, these results are more powerful because they are somewhat counterintuitive.

Table 7 shows our test for H4. The outcome (Strike, Swing) occurs 48.94 percent of the time on 3-2 counts compared to 37.66 of the time on 2-2 counts. The outcome (Ball, Take) occurs 21.78 percent of the time on 3-2 counts compared to 29.73 percent of the time on 2-2 counts. These results are significant at a 1-percent level, so we fail to reject H4.

In conclusion, our results from the pitch location model are mostly supportive of the Minimax Theorem. Our finding that batters swing more often at balls on 3-2 counts and thus often deprive themselves of an opportunity to get on base is particularly interesting. We will now move on to the pitch type model to test our last two hypotheses.

---

22 This is significant at a 5-percent level, but not at a 1-percent level. See Table 6 for details.
**PITCH TYPE MODEL**

In order to test H5 and H6, we need to restrict our data set. We restrict our data set to pitches on 0-0 counts to test H5, and pitches on all other counts to test H6. In both cases, various data issues make some additional pitches unusable. These issues are discussed in Table 4 of Appendix B. Pitches classified as “0-0, Usable,” which make up 25.76% of our data set, are used to test H5. Pitches classified as “Other, Usable,” which make up 73.30% of our data set, are used to test H6. We will now consider the reasons for restricting our data set in each case.

For H5, we restrict our data set to only pitches that begin an at-bat. This is to avoid including at-bats more than once in our regressions. We do not use the last pitch of the at-bat (as Kovash and Levitt (2009) did) in order to eliminate the possible introduction of selection bias. For example, off-speed pitches might be more difficult to hit than fastballs, but they might also be more difficult to throw for strikes. Using the last pitch of the at-bat would ignore the drawbacks of off-speed pitches not thrown in 3-ball counts, which would introduce a bias in favor of off-speed pitches. We would then expect a pitcher to consistently achieve worse results on fastballs that end an at-bat even if he were behaving optimally. Thus testing H5 with pitches that end an at-bat is not viable. On the other hand, if a pitcher achieves consistently worse results for first pitch fastballs than for first pitch off-speed pitches, he should strictly prefer first pitch off-speed pitches. Thus if we use the first pitch of the at-bat, we can include a pitch from every at-bat and still be confident that H5 should continue to hold.
For H6, we use only pitches that are not the first pitch of an at-bat. This allows us to consider the effect of the pitcher’s previous choice of pitch on his current choice. We choose to test serial independence only within individual at-bats. It is true that the pitcher’s choice on the first pitch of a particular at-bat should not depend on his choice on the last pitch of the previous at-bat. However, because payoffs differ between different pitcher-batter pairs, considering serial independence across at-bats would complicate the situation, making it more difficult to determine the reliability of any conclusion. Thus it is best if we restrict our data set to pitches which do not begin an at-bat.

We will now outline the tests for both H5 and H6. The general regression equations are presented in Figure 9 below.

\[
\begin{align*}
H5^* : \quad \text{Outcome}_{kt} &= \beta_1 \text{fastball}_{kt} + X_{kt} \alpha_1 + \delta_k + \varepsilon_{kt} \\
H6^* : \quad \text{fastball}_{kt} &= \beta_2 \text{fastball}_{k(t-1)} + Y_{kt} \alpha_2 + \lambda_k + u_{kt}
\end{align*}
\]

Figure 9

For both regression equations, \( k \) indexes a pitcher-batter pair and \( t \) indexes a pitch between the pair. \( \text{fastball} \) is a fastball indicator variable in both equations, while \( \delta \) and \( \lambda \) are pitcher-batter pair fixed effects and \( \varepsilon \) and \( u \) are errors terms. For H5\(^*\), we use either OPS (on-base plus slugging) or wOBA (weighted on-base average) for \( \text{Outcome} \).\(^{23}\) OPS was used by Kovash and Levitt (2009). However, OPS has been criticized for weighting certain outcomes too heavily. For this reason, we also use wOBA, which is a similar stat that weights outcomes based on their impact on run production.\(^{24}\) The choice of outcome variable does not substantially affect

\[^{23}\] The values for OPS and wOBA appear in Table 8 of Appendix B.

\[^{24}\] For common criticisms of OPS and an in-depth description of wOBA, see Beneventano, Berger, and Weinberg (2012).
our conclusions. The set of controls $X$ in $H_5^*$ includes indicator variables for the score (in terms of margin), inning, home team, baserunner configuration, and number of outs. The set of controls $Y$ in $H_6^*$ includes all the indicator variables in $X$ with the addition of indicator variables for the count. For both $H_5^*$ and $H_6^*$, we run some regressions with no fixed effects or controls, some with fixed effects but no controls, and some with both fixed effects and controls. Because the payoff matrix is different for each pitcher-batter pair and game situation, the models that include fixed effects and controls are most reliable.

In the case of $H_5^*$, we sometimes restrict our data set further by eliminating pitcher-batter pairs in which only one type of pitch is thrown on 0-0 counts. In such cases, we cannot be sure the pitcher is in fact playing a mixed strategy. Since the indifference requirement only holds for actions played with positive probability, a pitcher playing a pure strategy might not be indifferent between fastballs and off-speed pitches. If we eliminate all cases where the pitcher might be playing a pure strategy, we can be more certain that $H_5$ should hold. This restriction also makes our fixed effects regressions more informative, because each pair of our restricted data set gives observations for both pitch types.

For $H_6^*$, we use both a linear probability model and a logit model in testing. We also sometimes restrict our data set to only those pairs which contribute at least 27 usable observations. This number is chosen to eliminate all pitcher-batter pairs in which only one pitch type is thrown. Fixed effects models rely on within-group variation, so pairs using only one pitch type on usable pitches are not able to contribute to estimation. This could create a selection bias, since pairs that show
more frequent alternation have a greater impact on estimation. Eliminating only these pairs as we did in the previous case would cause the problem rather than solve it. We limit our data set based on number of usable observations, which will not lead to any selection bias. Since all pairs that contribute at least 27 usable observations\footnote{In the case of one pitcher-batter pair with 26 usable observations, the pitcher threw only fastballs.} use both fastballs and off-speed pitches, we can solve our problem by limiting our data set to only these pairs.

The results of our tests of H5 and H6 are reported in Table 9 and Table 10, respectively. In Table 9, 12 tests are presented. For each test, we include the estimated value of the coefficient on fastball, its standard error, the $p$-value, our choice of outcome variable, the number of observations used, and whether we include fixed effects and/or controls. In Table 10, 10 tests are presented. The information provided in Table 10 is the same as in Table 9 except no choice of outcome variable is involved and the odds ratio with its standard error is presented for all tests using a logit model. At the bottom of each table, we report which tests use the additional restrictions described above.

As we can see in Table 9, all of our tests fail to reject H5. The choice between OPS and wOBA has an impact on the estimators and standard errors, but neither choice provides evidence against H5 in any of our tests. It should be noted that Tests (5)-(6) and (11)-(12) provide the most reliable results because they include pitcher-batter fixed effects and controls for game situation. Thus empirical testing supports that pitchers on average achieve equivalent results from fastballs and off-speed pitches.

\footnote{In the case of one pitcher-batter pair with 26 usable observations, the pitcher threw only fastballs.}
In Table 10, we find evidence against H6. 8 of the 10 tests displayed in Table 10 reject H6 at a 1-percent level, with only Tests (7) and (9) failing to reject. In all but Tests (1) and (4), which do not use fixed effects or controls, we obtain a negative estimated coefficient on fastball. Although they use a smaller number of observations, Tests (8) and (10) provide the most reliable results, since they use a data set that does not suffer from selection bias and include both fixed effects and controls. Test (8), which comes from a linear probability model, indicates that the probability a pitcher throws a fastball is reduced by 2.9 percentage points if the previous pitch was a fastball. Additionally, Test (10), which comes from a logit model, indicates that a pitcher is 12.6% less likely to throw a fastball if the previous pitch was a fastball. Our results then strongly suggest that pitchers alternate too frequently between fastballs and off-speed pitches. Thus empirical testing reveals that pitchers do not exhibit serial independence, so we reject H6.

In our tests with the pitch type model, we find that pitchers on average achieve equivalent results with fastballs and off-speed pitches, in contrast to the findings of Kovash and Levitt. However, pitchers alternate between fastballs and off-speed pitches more than optimal behavior would suggest. We will briefly interpret these results in the next section.

CONCLUSION

SUMMARY OF RESULTS

In conclusion, we obtained results mostly supportive of the Minimax Theorem. Our four hypotheses from the pitch location model were confirmed by data, including those that were counterintuitive. From the pitch type model, H5 was
also confirmed through testing. However, we rejected the hypothesis of serial independence. Our results are thus in the same spirit as those obtained by Walker and Wooders (2001) in their study on tennis serves. Professional athletes seem to behave optimally except when it comes to avoiding patterns in repeated play.

How should we interpret these results? Perhaps the pitcher’s errors are too small for the batter to identify and exploit. We know people do not have a good intuitive understanding of randomness, so it is not hard to imagine players might have difficulty recognizing whether or not a sequence of choices is truly random. This interpretation indicates that players are not perfectly rational as we typically assume. However, our confirmation of H1-H5 indicates that the players’ strategies mimic true randomness closely enough that most conclusions of the Minimax Theorem still hold. Perhaps we should view the Minimax Theorem in this light: although people often fail to be perfectly rational, their behavior is rational enough that it largely conforms to the conclusions of the Minimax Theorem.

On the other hand, we might question whether pitchers should truly exhibit serial independence. Perhaps seeing a fastball makes the batter momentarily better at hitting fastballs. If this is the case, the pitcher would have to throw more off-speed pitches in order to make the batter indifferent. This example would give us exactly the results we received in our tests of H6. Thus maybe we should not view our rejection of H6 as evidence against the Minimax Theorem at all. Under this interpretation, this study provides nearly unqualified support of the Minimax Theorem.
There are a number of different issues future research should address. The hypotheses in this study, in particular H3, could be better analyzed with a larger data set.\textsuperscript{26} The models in this study could be developed further as well. More generally and perhaps most importantly, the issue of serial independence should be a focus of future studies on the Minimax Theorem. In order to determine the practical applicability of the Minimax Theorem, we must better understand how people attempt to mimic randomness and how they respond to it. Finally, the vast availability of statistics makes baseball ideal for research. Going forward, baseball should be prominently used in experiments on the Minimax Theorem.

\footnote{Kovash and Levitt used pitch data from the 2002-2006 Major League Baseball seasons, comprising a data set five times the size of the one used in this study.}
References

Bar-Hillel, Maya and Wagenaar, Willem A. “The Perception of Randomness.”

Beneventano, Philip, Paul D. Berger, and Bruce D. Weinberg. "Predicting Run
Production and Run Prevention in Baseball: The Impact of Sabermetrics."

Brown, James N. and Rosenthal, Robert W. “Testing the Minimax Hypothesis: A Re-
examination of O’Neill’s Game Experiment.” Econometrica, 58.5 (1990):
1065-1081.

Casti, J. L. Five Golden Rules: Great Theories of 20th-century Mathematics and Why

Chiappori, Pierre-Andre; Levitt, Steven and Groseclose, Timothy. “Testing Mixed
Strategy Equilibria When Players Are Heterogenous: The case of Penalty

Erev, Ido, and Alvin E. Roth. "Predicting how people play games: Reinforcement
learning in experimental games with unique, mixed strategy equilibria."

Hsu, Shih-Hsun, Chen-Ying Huang, and Cheng-Tao Tang. "Minimax play at


Appendix A

A.1

Appendix A.1 shows the formal derivation of Assumptions (5)-(8).

In the Models and Assumptions section, we determined that $\pi_{BC}$, $\pi_{BA}$, $\pi_{SC}$, and $\pi_{SA}$ will all be lower in a 3-2 count than in a 2-2 count. This gives us the following inequality:

(A.1.1) $\pi_{XY}^1 - \pi_{XY}^2 > 0 \ \forall X \in \{S,B\}, \forall Y \in \{A,C\}$

We next determined that the largest difference between 2-2 payoffs and 3-2 payoffs occurs in $\pi_{BC}$. This statement together with (A.1.1) gives us the following two inequalities:

(A.1.2) $\pi_{BC}^1 - \pi_{BC}^2 > \pi_{BA}^1 - \pi_{BA}^2$

(A.1.3) $\pi_{BC}^1 - \pi_{BC}^2 > \pi_{SC}^1 - \pi_{SC}^2$

We also determined that the smallest difference between 2-2 payoffs and 3-2 payoffs occurs in $\pi_{SA}$. Using (A.1.1), this gives us:

(A.1.4) $\pi_{SC}^1 - \pi_{SC}^2 > \pi_{SA}^1 - \pi_{SA}^2$

(A.1.5) $\pi_{BA}^1 - \pi_{BA}^2 > \pi_{SA}^1 - \pi_{SA}^2$

Either statement could produce the addition inequality below:

(A.1.6) $\pi_{BC}^1 - \pi_{BC}^2 > \pi_{SA}^1 - \pi_{SA}^2$

However, (A.1.6) does not help us formulate any useful assumptions.

By subtracting $\pi_{BC}^1$ and adding $\pi_{BA}^2$ in (A.1.2), we obtain:

Assumption (5) $\pi_{BA}^2 - \pi_{BC}^2 > \pi_{BA}^1 - \pi_{BC}^1$

By subtracting $\pi_{BC}^1$ and adding $\pi_{SC}^2$ in (A.1.3), we obtain:
Assumption (7) \( \pi_{SC}^2 - \pi_{BC}^2 > \pi_{SC}^1 - \pi_{BC}^1 \)

By subtracting \( \pi_{SA}^1 \) and adding \( \pi_{SC}^2 \) in (A.1.4), we obtain:

Assumption (6) \( \pi_{SC}^1 - \pi_{SA}^1 > \pi_{SC}^2 - \pi_{SA}^2 \)

By subtracting \( \pi_{SA}^1 \) and adding \( \pi_{BA}^2 \) in (A.1.5), we obtain:

Assumption (8) \( \pi_{BA}^1 - \pi_{SA}^1 > \pi_{BA}^2 - \pi_{SA}^2 \)

This completes the derivation of Assumptions (5)-(8).

A.2

Appendix A.2 shows the computation of the equilibrium solution for the pitch location model. It should be noted that this computation is nearly identical to the computation in Appendix A.3.

Let the pitcher’s strategy be \((p_S, p_B)\), where \(p_S\) and \(p_B\) are the probabilities that the pitcher plays S and B, respectively. Let the batter’s strategy be \((h_A, h_C)\), where \(h_A\) and \(h_C\) are the probabilities that the batter plays A and C, respectively. Since \(p_S + p_B = 1\) and \(h_A + h_C = 1\), these strategies can be rewritten as \((p_S, 1 - p_S)\) and \((h_A, 1 - h_A)\). Let \(U_K(X)\) be the expected payoff that player K (P for the pitcher and B for the batter) receives from playing the strategy X (S or B for the pitcher and A or C for the batter). Using the payoff matrix for the pitch location model (displayed above as Figure A-1), we obtain the following four expressions:

\[ (A.2.1) \quad U_P(S) = h_A \pi_{SA} + (1 - h_A) \pi_{SC} \]
Under the Minimax Theorem, both players must be indifferent between actions played with positive probability in equilibrium. Therefore, we obtain:

(A.2.5) \( U_p (S) = U_p (B) \)

(A.2.6) \( U_B (A) = U_B (C) \)

Combining the expressions (A.2.1), (A.2.2), and (A.2.5), we obtain:

(A.2.7) \( h_A \pi_{SA} + (1 - h_A) \pi_{SC} = h_A \pi_{BA} + (1 - h_A) \pi_{BC} \)

Combining the expressions (A.2.3), (A.2.4), and (A.2.6), we obtain:

(A.2.8) \( p_S \pi_{SA} + (1 - p_S) \pi_{BA} = p_S \pi_{SC} + (1 - p_S) \pi_{BC} \)

Both (A.2.7) and (A.2.8) can be simplified and factored to obtain:

(A.2.9) \( h_A (\pi_{SA} - \pi_{SC}) + \pi_{SC} = h_A (\pi_{BA} - \pi_{BC}) + \pi_{BC} \)

(A.2.10) \( p_S (\pi_{SA} - \pi_{BA}) + \pi_{BA} = p_S (\pi_{SC} - \pi_{BC}) + \pi_{BC} \)

Subtracting and dividing in (A.2.9) and (A.2.10) gives us:

(A.2.11) \( h_A = \frac{\pi_{SC} - \pi_{BC}}{\pi_{SC} - \pi_{BC} + \pi_{BA} - \pi_{SA}} \)

(A.2.12) \( p_S = \frac{\pi_{BA} - \pi_{BC}}{\pi_{BA} - \pi_{BC} + \pi_{SC} - \pi_{SA}} \)

These two expressions fully specify the equilibrium strategies for both players in the pitch location model.
A.3

Appendix A.3 shows the computation of the equilibrium solution for the pitch type model. It should be noted that this computation is nearly identical to the computation in Appendix A.2.

\[
\begin{array}{c|cc}
\text{Pitcher} & F & O \\
\hline
F & \pi_{FF} & \pi_{FO} \\
O & \pi_{OF} & \pi_{OO}
\end{array}
\]

Figure A-2

Let the pitcher’s strategy be \((p_F, p_O)\), where \(p_F\) and \(p_O\) are the probabilities that the pitcher plays F and O, respectively. Let the batter’s strategy be \((h_F, h_O)\), where \(h_F\) and \(h_O\) are the probabilities that the batter plays F and O, respectively. Since \(p_F + p_O = 1\) and \(h_F + h_O = 1\), these strategies can be rewritten as \((p_F, 1 - p_F)\) and \((h_F, 1 - h_F)\). Let \(U_K(X)\) be the expected payoff that player K (P for the pitcher and B for the batter) receives from playing the strategy X (F or O for either player). Using the payoff matrix for the pitch type model (displayed above as Figure A-2), we obtain the following four expressions:

\[(A.3.1)\quad U_P(F) = h_F\pi_{FF} + (1 - h_F)\pi_{FO}\]

\[(A.3.2)\quad U_P(O) = h_F\pi_{OF} + (1 - h_F)\pi_{OO}\]

\[(A.3.3)\quad U_B(F) = p_F\pi_{FF} + (1 - p_F)\pi_{OF}\]

\[(A.3.4)\quad U_B(O) = p_F\pi_{FO} + (1 - p_F)\pi_{OO}\]

Under the Minimax Theorem, both players must be indifferent between actions played with positive probability in equilibrium. Therefore, we obtain:

\[(A.3.5)\quad U_P(F) = U_P(O)\]
(A.3.6) \( U_B(F) = U_B(O) \)

Combining the expressions (A.3.1), (A.3.2), and (A.3.5), we obtain:

(A.3.7) \( h_F \pi_{FF} + (1 - h_F) \pi_{FO} = h_F \pi_{OF} + (1 - h_F) \pi_{OO} \)

Combining the expressions (A.3.3), (A.3.4), and (A.3.6), we obtain:

(A.3.8) \( p_F \pi_{FF} + (1 - p_F) \pi_{OF} = p_F \pi_{FO} + (1 - p_F) \pi_{OO} \)

Both (A.3.7) and (A.3.8) can be simplified and factored to obtain:

(A.3.9) \( h_F (\pi_{FF} - \pi_{FO}) + \pi_{FO} = h_F (\pi_{OF} - \pi_{OO}) + \pi_{OO} \)

(A.3.10) \( p_F (\pi_{FF} - \pi_{OF}) + \pi_{OF} = p_F (\pi_{FO} - \pi_{OO}) + \pi_{OO} \)

Subtracting and dividing in (A.3.9) and (A.3.10) gives us:

(A.3.11) \( h_F = \frac{\pi_{FO} - \pi_{OO}}{\pi_{FO} - \pi_{OO} + \pi_{OF} - \pi_{FF}} \)

(A.3.12) \( p_F = \frac{\pi_{OF} - \pi_{OO}}{\pi_{OF} - \pi_{OO} + \pi_{FO} - \pi_{FF}} \)

These two expressions fully specify the equilibrium strategies for both players in the pitch type model.

A.4

Appendix A.4 shows the formal justification of the claim that pitchers play “strike” more often on 3-2 counts than on 2-2 counts. This claim is the primary justification for H1. It should be noted that this computation is nearly identical to the computation in A.5.

Consider a fixed batter and pitcher. Let \( p_S^1 \) denote the pitcher’s equilibrium strategy on a 2-2 count and let \( p_S^2 \) denote the pitcher’s equilibrium strategy in an identical game situation except on a 3-2 count. Let \( \pi_{XY}^1 \) denote the payoff that results from
the pitcher playing X and the batter playing Y in a 2-2 count and let \( \pi_{XY} \) denote the payoff that results from the pitcher playing X and the batter playing Y in a 3-2 count (where X is S or B and Y is A or C). We can then write the pitcher’s equilibrium strategies in each count as follows:

\[(A.4.1)\quad \pi_S^1 = \frac{\pi_{BA}^1 - \pi_{BC}^1}{\pi_{BA}^1 - \pi_{BC}^1 + \pi_{SC}^1 - \pi_{SA}^1}\]

\[(A.4.2)\quad \pi_S^2 = \frac{\pi_{BA}^2 - \pi_{BC}^2}{\pi_{BA}^2 - \pi_{BC}^2 + \pi_{SC}^2 - \pi_{SA}^2}\]

Dividing in (A.4.1) and (A.4.2) gives us:

\[(A.4.3)\quad \pi_S^1 = \frac{1}{1 + \frac{\pi_{SC}^1 - \pi_{SA}^1}{\pi_{BA}^1 - \pi_{BC}^1}}\]

\[(A.4.4)\quad \pi_S^2 = \frac{1}{1 + \frac{\pi_{SC}^2 - \pi_{SA}^2}{\pi_{BA}^2 - \pi_{BC}^2}}\]

Assumptions (1), (2), (5), and (6) are recalled in Figure A-3 below:

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>( \pi_{BA} &gt; \pi_{BC} )</td>
</tr>
<tr>
<td>(2)</td>
<td>( \pi_{SC} &gt; \pi_{SA} )</td>
</tr>
<tr>
<td>(5)</td>
<td>( \pi_{BA}^2 - \pi_{BC}^2 &gt; \pi_{BA}^1 - \pi_{BC}^1 )</td>
</tr>
<tr>
<td>(6)</td>
<td>( \pi_{SC}^1 - \pi_{SA}^1 &gt; \pi_{SC}^2 - \pi_{SA}^2 )</td>
</tr>
</tbody>
</table>

Using Assumption (1), we obtain:

\[(A.4.5)\quad \pi_{BA}^1 - \pi_{BC}^1 > 0\]

\[(A.4.6)\quad \pi_{BA}^2 - \pi_{BC}^2 > 0\]

Using Assumption (2), we obtain:

\[(A.4.7)\quad \pi_{SC}^1 - \pi_{SA}^1 > 0\]
Using (A.4.5)-(A.4.8) and by combining Assumptions (5) and (6), we obtain:

\[(A.4.9) \ (\pi_{BA}^2 - \pi_{BC}^2)(\pi_{SC} - \pi_{SA}) > (\pi_{BA}^1 - \pi_{BC}^1)(\pi_{SC}^2 - \pi_{SA}^2) \]

Dividing in (A.4.9) gives us:

\[(A.4.10) \ \frac{\pi_{SC} - \pi_{SA}}{\pi_{BA} - \pi_{BC}} > \frac{\pi_{SC}^2 - \pi_{SA}^2}{\pi_{BA}^2 - \pi_{BC}^2} \]

Note that by (A.4.5)-(A.4.8):

\[(A.4.11) \ \frac{\pi_{SC} - \pi_{SA}}{\pi_{BA} - \pi_{BC}} > 0 \]

\[(A.4.12) \ \frac{\pi_{SC}^2 - \pi_{SA}^2}{\pi_{BA}^2 - \pi_{BC}^2} > 0 \]

The function \( f(t) = \frac{1}{1 + t} \) is strictly decreasing on \( [0, \infty) \)

Therefore, by (A.4.10), we obtain:

\[(A.4.13) \ f(\frac{\pi_{SC}^2 - \pi_{SA}^2}{\pi_{BA}^2 - \pi_{BC}^2}) > f(\frac{\pi_{SC} - \pi_{SA}}{\pi_{BA} - \pi_{BC}}) \]

However, by (A.4.3) and (A.4.4), \( p_s^1 = f(\frac{\pi_{SC} - \pi_{SA}}{\pi_{BA} - \pi_{BC}}) \) and \( p_s^2 = f(\frac{\pi_{SC}^2 - \pi_{SA}^2}{\pi_{BA}^2 - \pi_{BC}^2}) \)

Therefore, by (A.4.13), \( p_s^2 > p_s^1 \).

This shows that pitchers will play “strike” more frequently in equilibrium on 3-2 counts than on 2-2 counts.

A.5

Appendix A.5 shows the formal justification of the claim that batters play “aggressive” more often on 3-2 counts than on 2-2 counts. This claim is part of the justification of
H2 and H3. It should be noted that this computation is nearly identical to the computation in A.4.

Consider a fixed batter and pitcher. Let $h_1^A$ denote the batter’s equilibrium strategy on a 2-2 count and let $h_2^A$ denote the batter’s equilibrium strategy in an identical game situation except on a 3-2 count. Let $\pi_{XY}^1$ denote the payoff that results from the pitcher playing X and the batter playing Y in a 2-2 count and let $\pi_{XY}^2$ denote the payoff that results from the pitcher playing X and the batter playing Y in a 3-2 count (where X is S or B and Y is A or C). We can then write the batter’s equilibrium strategies in each count as follows:

(A.5.1) $h_1^A = \frac{\pi_{SC}^1 - \pi_{BC}^1}{\pi_{SC}^1 - \pi_{BC}^1 + \pi_{BA}^1 - \pi_{SA}^1}$

(A.5.2) $h_2^A = \frac{\pi_{SC}^2 - \pi_{BC}^2}{\pi_{SC}^2 - \pi_{BC}^2 + \pi_{BA}^2 - \pi_{SA}^2}$

Dividing in (A.5.1) and (A.5.2) gives us:

(A.5.3) $h_1^A = \frac{1}{1 + \frac{\pi_{BA}^1 - \pi_{SA}^1}{\pi_{SC}^1 - \pi_{BC}^1}}$

(A.5.4) $h_2^A = \frac{1}{1 + \frac{\pi_{BA}^2 - \pi_{SA}^2}{\pi_{SC}^2 - \pi_{BC}^2}}$

Assumptions (3), (4), (7), and (8) are recalled in Figure A-4 below:

<table>
<thead>
<tr>
<th>Assumption (3)</th>
<th>$\pi_{SC} &gt; \pi_{BC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption (4)</td>
<td>$\pi_{BA} &gt; \pi_{SA}$</td>
</tr>
<tr>
<td>Assumption (7)</td>
<td>$\pi_{SC}^2 - \pi_{BC}^2 &gt; \pi_{SC}^1 - \pi_{BC}^1$</td>
</tr>
<tr>
<td>Assumption (8)</td>
<td>$\pi_{BA}^1 - \pi_{SA}^1 &gt; \pi_{BA}^2 - \pi_{SA}^2$</td>
</tr>
</tbody>
</table>

Figure A-4
Using Assumption (3), we obtain:

(A.5.5) $\pi_{SC}^1 - \pi_{BC}^1 > 0$

(A.5.6) $\pi_{SC}^2 - \pi_{BC}^2 > 0$

Using Assumption (4), we obtain:

(A.5.7) $\pi_{BA}^1 - \pi_{SA}^1 > 0$

(A.5.8) $\pi_{BA}^2 - \pi_{SA}^2 > 0$

Using (A.5.5)-(A.5.8) and by combining Assumptions (7) and (8), we obtain:

(A.5.9) $(\pi_{SC}^2 - \pi_{BC}^2)(\pi_{BA}^1 - \pi_{SA}^1) > (\pi_{SC}^1 - \pi_{BC}^1)(\pi_{BA}^2 - \pi_{SA}^2)$

Dividing in (A.5.9) gives us:

(A.5.10) $\frac{\pi_{BA}^1 - \pi_{SA}^1}{\pi_{SC}^1 - \pi_{BC}^1} > \frac{\pi_{BA}^2 - \pi_{SA}^2}{\pi_{SC}^2 - \pi_{BC}^2}$

Note that by (A.5.5)-(A.5.8):

(A.5.11) $\frac{\pi_{BA}^1 - \pi_{SA}^1}{\pi_{SC}^1 - \pi_{BC}^1} > 0$

(A.5.12) $\frac{\pi_{BA}^2 - \pi_{SA}^2}{\pi_{SC}^2 - \pi_{BC}^2} > 0$

The function $f(t) = \frac{1}{1 + t}$ is strictly decreasing on $[0, \infty)$

Therefore, by (A.5.10), we obtain:

(A.5.13) $f\left(\frac{\pi_{BA}^2 - \pi_{SA}^2}{\pi_{SC}^2 - \pi_{BC}^2}\right) > f\left(\frac{\pi_{BA}^1 - \pi_{SA}^1}{\pi_{SC}^1 - \pi_{BC}^1}\right)$

However, by (A.5.3) and (A.5.4), $h_A^1 = f\left(\frac{\pi_{BA}^1 - \pi_{SA}^1}{\pi_{SC}^1 - \pi_{BC}^1}\right)$ and $h_A^2 = f\left(\frac{\pi_{BA}^2 - \pi_{SA}^2}{\pi_{SC}^2 - \pi_{BC}^2}\right)$

Therefore, by (A.5.13), $h_A^2 > h_A^1$. 

55
This shows that batters will play “aggressive” more frequently in equilibrium on 3-2 counts than on 2-2 counts.

A.6

Appendix A.6 shows that the batter swings more often the more the pitcher plays “strike” and the more the batter plays “aggressive.” This claim is part of the justification of H2 and H3.

Let $s_{xy}$ denote the probability that the batter swings when the pitcher plays $X$ (either S or B) and the batter plays $Y$ (either A or C). We assumed in the Hypotheses section that $s_{SA}$ is the largest of the four swing probabilities while $s_{BC}$ is the smallest. These assumptions can be formalized as follows:

(A.6.1) $s_{SA} > s_{SC}$

(A.6.2) $s_{SA} > s_{BA}$

(A.6.3) $s_{SC} > s_{BC}$

(A.6.4) $s_{BA} > s_{BC}$

(A.6.5) $s_{SA} > s_{BC}$

We can rewrite (A.6.1)-(A.6.4) as:

(A.6.6) $s_{SA} - s_{SC} > 0$

(A.6.7) $s_{SA} - s_{BA} > 0$

(A.6.8) $s_{SC} - s_{BC} > 0$

(A.6.9) $s_{BA} - s_{BC} > 0$
Let $q_{xy}$ denote the probability that the pitcher plays $X$ and the batter plays $Y$. As explained in the Hypotheses section, we can describe the probability that the batter swings by the following function:

\[(A.6.10) \quad f(q_{SA}, q_{BA}, q_{BC}) = q_{SA}s_{SA} + q_{SC}s_{SC} + q_{BA}s_{BA} + q_{BC}s_{BC}\]

We know that the strategies of the pitcher and batter are independent in equilibrium, so we obtain the following result:

\[(A.6.11) \quad q_{xy} = p_X h_Y \forall X \in \{S,B\}, \forall Y \in \{A,C\}\]

We can use (A.6.11) to rewrite (A.6.10) as the following function:

\[(A.6.12) \quad f(p_S, h_A) = p_SH_A s_{SA} + p_S(1-h_A)s_{SC} + p_B h_A s_{BA} + p_B(1-h_A)s_{BC}\]

Since $p_S + p_B = 1$ and $h_A + h_C = 1$, we can rewrite (A.6.12) as:

\[(A.6.13) \quad f(p_S, h_A) = p_S h_A s_{SA} + p_S(1-h_A)s_{SC} + (1-p_S)h_A s_{BA} + (1-p_S)(1-h_A)s_{BC}\]

We can simplify and factor (A.6.13) in two different ways to yield the two following functions:

\[(A.6.14) \quad f(p_S, h_A) = p_S(h_A(s_{SA} - s_{BA}) + (1-h_A)(s_{SC} - s_{BC})) + h_A s_{BA} + (1-h_A)s_{BC}\]

\[(A.6.15) \quad f(p_S, h_A) = h_A(p_S(s_{SA} - s_{SC}) + (1-p_S)(s_{BA} - s_{BC})) + p_S s_{SC} + (1-p_S)s_{BC}\]

By taking the partial derivative with respect to $p_S$ in (A.6.14), we obtain:

\[(A.6.16) \quad \frac{\partial f}{\partial p_S}(p_S, h_A) = h_A(s_{SA} - s_{BA}) + (1-h_A)(s_{SC} - s_{BC})\]

By taking the partial derivative with respect to $h_A$ in (A.6.15), we obtain:

\[(A.6.17) \quad \frac{\partial f}{\partial h_A}(p_S, h_A) = p_S(s_{SA} - s_{SC}) + (1-p_S)(s_{BA} - s_{BC})\]

Since $p_S$ and $h_A$ are probabilities, the following hold:

\[(A.6.18) \quad p_S \geq 0\]
\( (A.6.19) \ 1 - p_s \geq 0 \)

\( (A.6.20) \ h_A \geq 0 \)

\( (A.6.21) \ 1 - h_A \geq 0 \)

By \((A.6.7)\), \((A.6.8)\), \((A.6.18)-(A.6.21)\), and \((A.6.16)\), we obtain:

\( (A.6.22) \ \frac{\partial f}{\partial p_s}(p_s,h_A) > 0 \)

By \((A.6.6)\), \((A.6.9)\), \((A.6.18)-(A.6.21)\), and \((A.6.17)\), we obtain:

\( (A.6.23) \ \frac{\partial f}{\partial h_A}(p_s,h_A) > 0 \)

Therefore, \( f(p_s,h_A) \) is strictly increasing in both \( p_s \) and \( h_A \).

Therefore, the probability that the batter swings is greater the more the pitcher plays “strike” and the batter plays “aggressive”.

This fact can be used to justify H2 and H3.
Appendix B

Table 1

<table>
<thead>
<tr>
<th>Pitch Location</th>
<th>Pitches</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inside the Zone</td>
<td>311,262</td>
<td>44.13</td>
</tr>
<tr>
<td>Outside the Zone</td>
<td>394,059</td>
<td>55.87</td>
</tr>
<tr>
<td>In The Strike Zone</td>
<td>311,262</td>
<td>44.13</td>
</tr>
<tr>
<td>Outside</td>
<td>129,006</td>
<td>18.29</td>
</tr>
<tr>
<td>High</td>
<td>75,045</td>
<td>10.64</td>
</tr>
<tr>
<td>Inside</td>
<td>62,184</td>
<td>8.82</td>
</tr>
<tr>
<td>Low</td>
<td>38,883</td>
<td>5.51</td>
</tr>
<tr>
<td>High Outside</td>
<td>38,657</td>
<td>5.48</td>
</tr>
<tr>
<td>Low Outside</td>
<td>20,316</td>
<td>2.88</td>
</tr>
<tr>
<td>High Inside</td>
<td>18,350</td>
<td>2.60</td>
</tr>
<tr>
<td>Low Inside</td>
<td>11,618</td>
<td>1.65</td>
</tr>
<tr>
<td>Total</td>
<td>705,321</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Pitch Type</th>
<th>Pitches</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fastball</td>
<td>444,162</td>
<td>63.35</td>
</tr>
<tr>
<td>Off-Speed</td>
<td>256,949</td>
<td>36.65</td>
</tr>
<tr>
<td>Fastball</td>
<td>403,823</td>
<td>57.60</td>
</tr>
<tr>
<td>Cut Fastball</td>
<td>40,339</td>
<td>5.75</td>
</tr>
<tr>
<td>Slider</td>
<td>100,984</td>
<td>14.40</td>
</tr>
<tr>
<td>Curveball</td>
<td>72,589</td>
<td>10.35</td>
</tr>
<tr>
<td>Changeup</td>
<td>68,957</td>
<td>9.84</td>
</tr>
<tr>
<td>Split Finger</td>
<td>11,472</td>
<td>1.64</td>
</tr>
<tr>
<td>Knuckleball</td>
<td>2,862</td>
<td>0.41</td>
</tr>
<tr>
<td>Screwball</td>
<td>85</td>
<td>0.01</td>
</tr>
<tr>
<td>Total</td>
<td>701,111</td>
<td>100.00</td>
</tr>
</tbody>
</table>

4,210 pitches (0.60% of the data set) were labeled “No Video.”
### Table 3

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pitches</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strike/Ball</td>
<td>569,693</td>
<td>80.77</td>
</tr>
<tr>
<td>In Play</td>
<td>135,628</td>
<td>19.23</td>
</tr>
<tr>
<td>Strike</td>
<td>313,403</td>
<td>44.43</td>
</tr>
<tr>
<td>Ball</td>
<td>256,290</td>
<td>36.34</td>
</tr>
<tr>
<td>In Play, Out</td>
<td>93,565</td>
<td>13.27</td>
</tr>
<tr>
<td>In Play, No Out</td>
<td>42,063</td>
<td>5.96</td>
</tr>
<tr>
<td>Total</td>
<td>705,321</td>
<td>100.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pitches</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swing</td>
<td>324,092</td>
<td>45.95</td>
</tr>
<tr>
<td>Take</td>
<td>381,229</td>
<td>54.05</td>
</tr>
<tr>
<td>Inside the Zone, Swing</td>
<td>203,388</td>
<td>28.84</td>
</tr>
<tr>
<td>Outside the Zone, Swing</td>
<td>120,704</td>
<td>17.11</td>
</tr>
<tr>
<td>Inside the Zone, Take</td>
<td>107,874</td>
<td>15.29</td>
</tr>
<tr>
<td>Outside the Zone, Take</td>
<td>273,355</td>
<td>38.76</td>
</tr>
<tr>
<td>Total</td>
<td>705,321</td>
<td>100.00</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Subset of Data</th>
<th>Pitches</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-2/3-2</td>
<td>87,604</td>
<td>12.42</td>
</tr>
<tr>
<td>Other</td>
<td>617,717</td>
<td>87.58</td>
</tr>
<tr>
<td>2-2</td>
<td>55,158</td>
<td>7.82</td>
</tr>
<tr>
<td>3-2</td>
<td>32,446</td>
<td>4.60</td>
</tr>
<tr>
<td>Other</td>
<td>617,717</td>
<td>87.58</td>
</tr>
<tr>
<td>Total</td>
<td>705,321</td>
<td>100.00</td>
</tr>
<tr>
<td>0-0</td>
<td>184,609</td>
<td>26.17</td>
</tr>
<tr>
<td>Other</td>
<td>520,712</td>
<td>73.83</td>
</tr>
<tr>
<td>0-0, Usable</td>
<td>181,680</td>
<td>25.76</td>
</tr>
<tr>
<td>0-0, Not Usable</td>
<td>2,929</td>
<td>0.42</td>
</tr>
<tr>
<td>Other, Usable</td>
<td>516,998</td>
<td>73.30</td>
</tr>
<tr>
<td>Other, Not Usable</td>
<td>3,714</td>
<td>0.53</td>
</tr>
<tr>
<td>Total</td>
<td>705,321</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Notes: 0-0 pitches are classified as “Not Usable” if the pitch type is given as “No Video” (1,037 pitches) or if the at-bat result is given as “Sacrifice” (1,377 pitches), “Sacrifice With Error” (75 pitches), “Interference” (23 pitches), or “NULL” (430 pitches). 0-0 pitches with an at-bat result of “Intentionally Walked” (1,055 pitches) can be usable with OPS as a dependent variable, but not with wOBA. Pitches in counts besides 0-0 are classified as “Not Usable” if the pitch type is given as “No Video” (3,173 pitches), or if the pitch type for the previous pitch is given as “No Video” (3,120 pitches), or because of miscellaneous errors in the data (23 pitches).
### Table 5

<table>
<thead>
<tr>
<th>Count</th>
<th>In the Strike Zone</th>
<th>Total</th>
<th>Strike Zone %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-2</td>
<td>23,071</td>
<td>55,158</td>
<td>41.83</td>
</tr>
<tr>
<td>3-2</td>
<td>17,183</td>
<td>32,446</td>
<td>52.96</td>
</tr>
<tr>
<td>Total</td>
<td>40,254</td>
<td>87,604</td>
<td>45.95</td>
</tr>
</tbody>
</table>

\( H_0: p(\text{Strike}|3-2) = p(\text{Strike}|2-2) \)
\( H_a: p(\text{Strike}|3-2) > p(\text{Strike}|2-2) \)

\( p \)-value for one-sided t-test: .000***

***Significant at a 1-percent level

### Table 6

<table>
<thead>
<tr>
<th>Pitch Location</th>
<th>2-2 Counts</th>
<th>3-2 Counts</th>
<th>2-2 Counts</th>
<th>3-2 Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Swings</td>
<td>Total</td>
<td>%</td>
<td>Swings</td>
</tr>
<tr>
<td>Total</td>
<td>36,463</td>
<td>55,158</td>
<td>66.11</td>
<td>24,076</td>
</tr>
<tr>
<td>Inside the Zone</td>
<td>20,774</td>
<td>23,071</td>
<td>90.04</td>
<td>15,880</td>
</tr>
<tr>
<td>Outside the Zone</td>
<td>15,689</td>
<td>32,087</td>
<td>48.90</td>
<td>8,196</td>
</tr>
<tr>
<td>In The Strike Zone</td>
<td>20,774</td>
<td>23,071</td>
<td>90.04</td>
<td>15,880</td>
</tr>
<tr>
<td>Outside</td>
<td>5,355</td>
<td>9,832</td>
<td>54.47</td>
<td>3,121</td>
</tr>
<tr>
<td>High</td>
<td>3,569</td>
<td>6,610</td>
<td>54.99</td>
<td>1,602</td>
</tr>
<tr>
<td>Inside</td>
<td>2,801</td>
<td>5,324</td>
<td>52.61</td>
<td>1,509</td>
</tr>
<tr>
<td>Low</td>
<td>1,828</td>
<td>2,982</td>
<td>61.30</td>
<td>1,066</td>
</tr>
<tr>
<td>High Outside</td>
<td>1,002</td>
<td>3,346</td>
<td>29.95</td>
<td>367</td>
</tr>
<tr>
<td>High Inside</td>
<td>582</td>
<td>1,901</td>
<td>30.62</td>
<td>214</td>
</tr>
<tr>
<td>Low Outside</td>
<td>284</td>
<td>1,328</td>
<td>21.39</td>
<td>188</td>
</tr>
<tr>
<td>Low Inside</td>
<td>268</td>
<td>764</td>
<td>35.08</td>
<td>129</td>
</tr>
</tbody>
</table>

\( H_0: p(\text{Swing}|3-2) = p(\text{Swing}|2-2) \)
\( H_a: p(\text{Swing}|3-2) > p(\text{Swing}|2-2) \)

\( p \)-values are for a one-sided t-test

**Significant at a 5-percent level

***Significant at a 1-percent level
### Table 7

<table>
<thead>
<tr>
<th>Count</th>
<th>(S,S)</th>
<th>(B,T)</th>
<th>All Outcomes</th>
<th>(S,S)%</th>
<th>(B,T)%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-2</td>
<td>20,774</td>
<td>16,398</td>
<td>55,158</td>
<td>37.66</td>
<td>29.73</td>
</tr>
<tr>
<td>3-2</td>
<td>15,880</td>
<td>7,067</td>
<td>32,446</td>
<td>48.94</td>
<td>21.78</td>
</tr>
<tr>
<td>2-2/3-2</td>
<td>36,654</td>
<td>23,465</td>
<td>87,604</td>
<td>41.84</td>
<td>26.79</td>
</tr>
</tbody>
</table>

Notes: (S,S) denotes the outcome where the pitcher throws the ball in the strike zone and the batter swings. (B,T) denotes the outcome where the pitcher throws the ball outside the strike zone and the batter does not swing. (S,S)% and (B,T)% are the frequencies with which (S,S) and (B,T) occur, respectively.

- $H_0$: $p((S,S)|3-2)=p((S,S)|2-2)$
- $H_A$: $p((S,S)|3-2)>p((S,S)|2-2)$
- *p*-value for one-sided t-test for (S,S): .000***

- $H_0$: $p((B,T)|3-2)=p((B,T)|2-2)$
- $H_A$: $p((B,T)|3-2)<p((B,T)|2-2)$
- *p*-value for one-sided t-test for (B,T): .000***

***Significant at a 1-percent level

### Table 8

<table>
<thead>
<tr>
<th>Metric</th>
<th>Out</th>
<th>INT</th>
<th>SAC</th>
<th>IW</th>
<th>Walk</th>
<th>HBP</th>
<th>Single</th>
<th>Double</th>
<th>Triple</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPS</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>wOBA</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>.691</td>
<td>.722</td>
<td>.884</td>
<td>1.257</td>
<td>1.593</td>
<td>2.058</td>
<td></td>
</tr>
</tbody>
</table>

Notes: "Out" includes all outcomes where the batter fails to reach base safely except sacrifice bunts. "Out" also includes fielder’s choices and errors (except errors that occur on sacrifice bunts). "INT" denotes defensive interference. "SAC" includes successful sacrifice bunts and sacrifice bunts with errors. "IW" denotes an intentional walk, while "Walk" denotes an unintentional walk. "HBP" denotes a hit by pitch and "HR" denotes a home run. The rest of the categories are straightforward. All outcomes in the data besides those classified as "NULL" can be placed in one of these categories. The values for OPS and wOBA come from FanGraphs, a website reference by Beneventano, Berger, and Weinberg (2012).
<table>
<thead>
<tr>
<th>Test</th>
<th>( \beta_i )</th>
<th>( p)-value</th>
<th>Outcome Variable</th>
<th>Fixed Effects</th>
<th>Controls</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.002 (.006)</td>
<td>.637</td>
<td>OPS</td>
<td>No</td>
<td>No</td>
<td>181680</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.003 (.003)</td>
<td>.264</td>
<td>wOBA</td>
<td>No</td>
<td>No</td>
<td>181422</td>
</tr>
<tr>
<td>(3)</td>
<td>0.002 (.008)</td>
<td>.784</td>
<td>OPS</td>
<td>Yes</td>
<td>No</td>
<td>181680</td>
</tr>
<tr>
<td>(4)</td>
<td>-0.002 (.003)</td>
<td>.651</td>
<td>wOBA</td>
<td>Yes</td>
<td>No</td>
<td>181422</td>
</tr>
<tr>
<td>(5)</td>
<td>0.009 (.008)</td>
<td>.252</td>
<td>OPS</td>
<td>Yes</td>
<td>Yes</td>
<td>181680</td>
</tr>
<tr>
<td>(6)</td>
<td>0.002 (.004)</td>
<td>.592</td>
<td>wOBA</td>
<td>Yes</td>
<td>Yes</td>
<td>181422</td>
</tr>
<tr>
<td>(7)</td>
<td>0.005 (.008)</td>
<td>.503</td>
<td>OPS</td>
<td>No</td>
<td>No</td>
<td>104396</td>
</tr>
<tr>
<td>(8)</td>
<td>-0.000 (.003)</td>
<td>.957</td>
<td>wOBA</td>
<td>No</td>
<td>No</td>
<td>104249</td>
</tr>
<tr>
<td>(9)</td>
<td>0.002 (.008)</td>
<td>.786</td>
<td>OPS</td>
<td>Yes</td>
<td>No</td>
<td>104396</td>
</tr>
<tr>
<td>(10)</td>
<td>-0.002 (.003)</td>
<td>.655</td>
<td>wOBA</td>
<td>Yes</td>
<td>No</td>
<td>104249</td>
</tr>
<tr>
<td>(11)</td>
<td>.008 (.008)</td>
<td>.323</td>
<td>OPS</td>
<td>Yes</td>
<td>Yes</td>
<td>104396</td>
</tr>
<tr>
<td>(12)</td>
<td>.001 (.004)</td>
<td>.696</td>
<td>wOBA</td>
<td>Yes</td>
<td>Yes</td>
<td>104249</td>
</tr>
</tbody>
</table>

**Notes:** This table displays the results for the regression of an outcome variable on a fastball indicator variable. Tests (1)-(6) are conducted with all usable pitches leading off an at-bat (see Table 4). Tests (7)-(12) are conducted using only pairs in which both fastballs and off-speed pitches were thrown. The outcome variable is either OPS or wOBA. Some tests use pitcher-batter fixed effects. Some tests use a set of controls including indicator variables for the score (in terms of margin for the batting team), inning, home team, baserunner situation, and number of outs.

*Rejection at a 10-percent level*
<table>
<thead>
<tr>
<th>Test</th>
<th>$\beta_2$</th>
<th>Odds Ratio</th>
<th>$p$-value</th>
<th>Fixed Effects</th>
<th>Controls</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>.093 (.001)</td>
<td>N/A</td>
<td>.000***</td>
<td>No</td>
<td>No</td>
<td>516998</td>
</tr>
<tr>
<td>(2)</td>
<td>-.089 (.002)</td>
<td>N/A</td>
<td>.000***</td>
<td>Yes</td>
<td>No</td>
<td>516998</td>
</tr>
<tr>
<td>(3)</td>
<td>-.112 (.001)</td>
<td>N/A</td>
<td>.000***</td>
<td>Yes</td>
<td>Yes</td>
<td>516998</td>
</tr>
<tr>
<td>(4)</td>
<td>.387 (.006)</td>
<td>1.473 (.009)</td>
<td>.000***</td>
<td>No</td>
<td>No</td>
<td>516998</td>
</tr>
<tr>
<td>(5)</td>
<td>-.383 (.007)</td>
<td>.682 (.005)</td>
<td>.000***</td>
<td>Yes</td>
<td>No</td>
<td>462931</td>
</tr>
<tr>
<td>(6)</td>
<td>-.507 (.007)</td>
<td>.603 (.004)</td>
<td>.000***</td>
<td>Yes</td>
<td>Yes</td>
<td>462931</td>
</tr>
<tr>
<td>(7)</td>
<td>-.003 (.005)</td>
<td>N/A</td>
<td>.476</td>
<td>Yes</td>
<td>No</td>
<td>47620</td>
</tr>
<tr>
<td>(8)</td>
<td>-.029 (.005)</td>
<td>N/A</td>
<td>.000***</td>
<td>Yes</td>
<td>Yes</td>
<td>47620</td>
</tr>
<tr>
<td>(9)</td>
<td>-.014 (.021)</td>
<td>.986 (.020)</td>
<td>.483</td>
<td>Yes</td>
<td>No</td>
<td>47620</td>
</tr>
<tr>
<td>(10)</td>
<td>-.134 (.021)</td>
<td>.874 (.019)</td>
<td>.000***</td>
<td>Yes</td>
<td>Yes</td>
<td>47620</td>
</tr>
</tbody>
</table>

*Notes:* This table displays the results for the regression of a fastball indicator variable on a once-lagged fastball indicator variable. Tests (1)-(6) are conducted with all usable data (see Table 4), while the data set analyzed in Tests (7)-(10) is restricted to only those pairs with at least 27 usable observations. Tests (1)-(3) and (7)-(8) are conducted using a linear probability model. Tests (4)-(6) and (9)-(10) are conducted using a logit model. For all tests using a logit model, the odds ratio is reported in addition to the estimated value of $\beta$ (which is the natural log of the odds ratio). Tests using a logit model with fixed effects drop all pairs with no variation in fastball. Some tests use pitcher-batter fixed effects. Some tests use a set of controls including indicator variables for the score (in terms of margin for the batting team), inning, home team, baserunner situation, number of outs, and count.  

***Rejection at the 1-percent level