Discussion Papers In Economics And Business

Reference Dependence and Monetary Incentives:

Evidence from Major League Baseball

Reio Tanji

Discussion Paper 20-23

March 2021

Graduate School of Economics
Osaka University, Toyonaka, Osaka 560-0043, JAPAN

Reference Dependence and Monetary Incentives: Evidence from Major League Baseball*

Reio Tanji†

Abstract

This paper discusses what determines the reference point in decision making, using an empirical dataset of performance stats in professional baseball. Previous literature has argued that some round-numbers may work as such points, and as a result, bunching occurs around these numbers in the distribution of the target outcomes. On the other hand, in the setting of workplace, the outcomes are observable both for the worker (players) and the evaluator (managers). This paper shows that this bunching do NOT occur from the structure of the contracts, or how the managers evaluate the players. Bunching seems to stem from the reference-point dependence of the workers themselves, and so to avoid this economically inefficient behavior, we have to design contracts that incentivize players to do so.

JEL Classification: D91, J01

keywords: reference dependence, round-number effect, bunching, monetary incentive, sports

* The author would like to thank Professor Masaru Sasaki from Osaka University, our supervisor and the chief examiner of the master's thesis. The author is grateful for the advice of Professor Fumio Ohtake from Osaka University, our sub-chief supervisors, and Associate professor Sohtaro Katsumata from Osaka University, our sub-chief supervisors. The author is also grateful for the comments of the seminar participants at Ohtake-Sasaki Seminar at Osaka University, Graduate School of Economics.

† Osaka University, Graduate Shcool of Economics. Email: u280537h@ecs.osaka-u.ac.jp

1 Introduction

Reference-point dependence is one of the most important concepts to evaluate outcomes, and it affects agents' economic behavior. Classical economic models assume that economic agents evaluate their choices/prospects according to the absolute value of the (expected) return. On the other hand, Tversky and Kahneman (1992) introduced the behavioral assumption: agents with reference-point dependent references consider outcomes by the relative value to some target value of the outcomes, or reference points. That is, the agents have the feeling of gain or loss by the extent to which the current outcome deviates from the target. For example, workers feel happy if her/his wage goes up to \$15 per hour, but unhappy if it goes down to \$15, although the absolute value of \$15 remains the same. In this case, the previous wage works as the reference point and affects her/his utility.

In this paper, we explore whether there are bunching in the salary of the player whose performance stats are above round-numbers, using data on individual performance stats in Major League Baseball (MLB).

Prospect theory consists of two main characteristics: one is the probability weighting function, and the other is the reference-point dependence, mentioned above. They allow us to interpret phenomena that cannot be explained by the traditional microeconomic theory. To explore the mechanism, we observe that much following research has been conducted in field and laboratory settings.

Reference dependence also is observed in the behavior of athletes. Pope and Schweizer (2011) found that professional golf players regarded "par," the standard number of shots determined according to the difficulty of each hole, as reference points. Additionally, Allen et al. (2016) argued that marathon runners adjusted their finish times just before the round number times (just three or four hours), the reference points. Similarly, Pope and Simonsohn (2011) showed the existence of the reference point in the MLB, a professional baseball league of the U.S..

MLB position players evaluate themselves by the stats of their performance. Moreover, they seem to have some reference points in their self-evaluation, for example, about their batting performance stats: .300 of batting average. Pope and Simonsohn (2011) have shown that there exists bunching just above .300 of the distribution of these stats.

Pope and Simonsohn (2011) differs from the former two papers. Golfers or marathon runners do not receive any monetary incentives to achieve these goals, but MLB players receive monetary rewards determined according to their performance.

Consider the case of professional golf players. Golf is essentially a competition of the total number of shots they needed to finish the whole tour, regardless of that of every single hole, or whether she or he saves par or not in the hole. The final rank of order is determined according to the total number of shots, and those with better scores are rewarded. Then, there appears a question as to what if there is some monetary incentive to make an effort to save par. Suppose when every time she or he saves par in each hole, she or he can get some additional bonus separated from their total score. In this case, then, making an effort to save par can be interpreted as a sufficiently “rational” choice for the player, although the observed behavior itself appears to be evidence of the reference dependence. However, there usually does not exist any additional monetary rewards such as “Save-Par-Bonus.”

On the contrary, in MLB, it might not be sufficient to prove that the bunching is caused by reference-dependent utility function about their performance stats: team managers may assign some monetary incentives for the players to adjust their aspiration level to meet those points, as we described in the example above.

Our most important contribution of this paper is to reveal the observed behavior of MLB players to be in fact reference dependence. There are two things to be mentioned. First, we found the evidence for the manipulation of the performance stats around the round numbers, by using McCrary (2008)’s method to test manipulation. We additionally include analysis of not only .300 of batting average but also other points of batting-average and other stats, such as homerun or stolen-bases. Second, we explored whether this observed manipulation was truly driven by the preference for reference dependence of the players. We applied a regression analysis using the data of the players’ annual salary.

Our paper found three results. First, our examination for manipulation supported Pope and Simonsohn (2011). We observed there existed seemingly reference point dependent behavior, where .300 of batting-average worked as a reference point. Similar results were obtained about other round numbers of batting-average, and other batting stats such as on-base percentage or homerun. Second, we found that as a whole, there did not exist any monetary incentive for them: for their fixed part of the salary contract, we conclude that their monetary rewards are continuous at just above each performance stats, such as .300 of batting-average. That is, they behave as they consider these round numbers as reference points, even though they do not receive any additional payment by achieving them. Furthermore, we reinforce our results from some alternative interpretations about other types of monetary incentives: the part of the incentivized contract, and relation with contract length. Finally, there exist serial changes in the players’ stats manipulation. .250 of batting-average was not considered as a reference point in recent years, while 20 of homerun was only for the recent players. Among them, .300 of batting-average seems to be a solid benchmark for the players.

Through the results above, this paper cotributes to the studies that argues how to design the efficient contract. As the same as other workplace, players (workers) are evaluated by their managers (evaluators), based on some performance outcomes. Observed bunching is economically inefficient behavior for the managers, since reaching the round-numbers does not raise the expected winning-percentage of the teams (firm's productivity). On the other hand, the possible resource of this misallocation is not only the players themselves. In practice, paying attention to which one (or other possible mechanisms) affects to the observed inefficient behavior and adequately motivating individuals, such as designing contracts will enhance the workers productivity.

This paper proceeds as follows. In Section 2, we review some literature and verify the standpoint of my paper. Section 3 describes the data we availed. Section 4 presents the theoretical framework and empirical specification, and make some conjecture. Section 5 shows the results of the analysis. Discussion about some alternative interpretation and non-statistical data are included in Section 6. Finally, Section 7 provides concluding remarks.

2 Literature Review

Tversky and Kahneman (1992) mentioned reference point dependence as one of the two distinct respects of their prospect theory. The most primitive form of reference-point dependent utility function is:

$$u(x|r) = \begin{cases} x - r & \text{if } x \geq r \\ \lambda(x - r) & \text{if } x < r \end{cases}$$

where x denotes a certain outcome, and r is one of the reference points (Figure 1). This agent evaluates the outcome by a difference from the reference point. Besides, they assume "loss-aversion" of the individual, or $\lambda > 1$. The size of the disutility incurred when an individual loses a certain amount of outcome is larger than the size of utility gained when she or he receives the same amount of outcome. "Diminishing sensitivity," which is concave in the phase of gain and convex in that of loss is an advanced form of this specification (Figure 2).

Diecidue and Van de Ven (2008)'s "aspiration level" model added discontinuity assumption: that is, a utility function that "jumps" at the reference point (Figure 3). When there exists a jump in their utility function, then individuals try to manipulate their outcome level, paying additional cost which is not incorporated into the model with the standard continuous utility function. As a result, excess mass or bunching around or just above the reference point arises. We discuss the required functional assumptions of them in Section 4.1.

Individuals with such reference-dependent utility aim to put their effort to achieve their internal target or reference point. There is several empirical literature that specifies the existence of reference dependence in the field or lab experimented studies. Farber (2008) applied this model to the labor supply of New York cab drivers to show that as soon as they reached daily target sales, they stopped working, even when they reached it early. Jones (2018) analyzed the system of American tax payment. He showed that individuals tried to manipulate their real payment by substituting it by donation or other charitable action, and especially did so when facing losses. This observation is also caused by the feeling of loss-aversion, with the reference point of the zero-payment threshold.

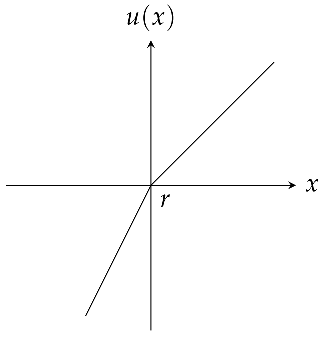

Figure 1: primitive gain-loss function

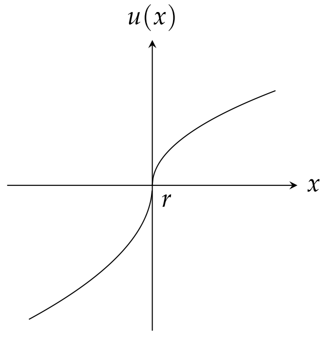

Figure 2: diminishing sensitivity

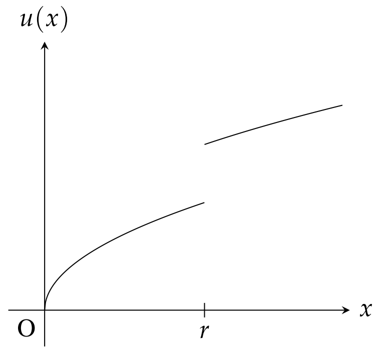

Figure 3: jump at the reference point

Reference dependence also arises in cases of sports. One of the most well-known papers among them is Pope and Schweizer (2011). They obtained the data of professional golf players, to show that in each hole, players behave as they take "par" as the reference point. Specifically, the probability of succeeding in their putts was significantly higher when the putt was one to save par than when it was one to get "eagle" or "birdie." Similarly, Allen *et al* (2016) specified the existence of reference point dependence of marathon runners, using data of the finish time of runners in an enormous number of races in the United States. In this case, the distribution of finishing time has an excess mass around every 30 minutes. Note that these cases are common in that the outcomes themselves are nonmonetary ones, and even if they achieve their internal goals, they do not receive any additional monetary reward for their success. Professional golf players are awarded according to the total number of shots through the whole tour, not to the number of pars they saved.

Pope and Simonsohn (2011) mention a seemingly similar case. They picked up three empirical pieces of evidence of round numbers that were considered as reference points: SAT (a standardized test for college admission in the United States) scores, laboratory experiments, and baseball. In their section of baseball, they picked up the evidence of Major League Baseball (MLB) players. They argue that the players have reference-dependent preferences for evaluating themselves in their performance stats: batting-average (AVG). According to their paper, the position players (batters) pay attention to their batting-average (AVG) especially to finish each season with their batting average of just above .300. They obtained MLB season individual AVG data from 1975 to 2008 and observed position players (= players except for pitchers) with at least 200 at-bats in each season. Then, they found that their distribution of the batting-average had excess mass just above .300, which revealed the existence of manipulation there. Furthermore, they found that players with batting-average of just below .300 were more likely to hit a base-hit and less likely to get a base-on-balls. Both base-hits and base-on-balls avoid the batter from being gotten out, so for the team which he belongs to, base-on-balls also has important value to win the game. However, batting-average does not count base-on-balls as an element to raise the number (For the definition of performance stats, see Appendix), so they prefer getting hit to base-on-balls.

There is similar behavior to the cases of Allen *et al* (2016): bunching. However, one important thing we have to take care of is that they are professional athletes, and so there exists a procedure of the contract between the player and the team managers: those who evaluate the player. They propose contracts to the players for each season before it starts. In other words, the observed excess mass may be derived by their monetary value function, not by the preference of themselves. We will distinguish them, which is the main determining factor of the bunching behavior of this paper.

Pope and Simonsohn (2011) conducted analysis only for batting-average. To reinforce their research, below are our following topics. We first search for the round numbers of various batting stats manipulated by the players. Then, we test if there exist any monetary incentives for the players. The team managers and the players agree to the contract that sets fixed additional bonuses which are paid when certain performance stats reaches the point. For the players whose stats are around the cutoff points, making a discontinuously large effort to succeed in bunching can be interpreted as an economically rational choice, under the above contract. For example, there should be little difference in the skills as a player between players with .300 of batting-average and .299. Thus, rational team managers contract them almost the same offer to these two players, controlling other factors than the batting-average.

Therefore, two possibilities are to be concerned. One is that team managers offer contracts that make players make additional efforts to achieve the goals. In this case, we can conclude that the managers show reference dependence about evaluating players. The other is that players have reference-dependent preferences, which leads to observed excess mass around the cutoff points. In this case, contracts do not bunch at any round-numbers.

Research about performance and monetary rewards using Major League Baseball has been conducted in the field of labor economics. One of the first papers is published by Scully (1974). In the next section, we describe the data used in our paper.

3 Data Description

To make empirical research, we need information about the players' performance, contracts, and other details. Then, we obtained panel data that contain this specific information from some open data-source. Each sample consists of stats of a player at the end of a single season. We explain the specific information about the dataset below.

First, Performance data are obtained from baseball fan website: *FanGraphs* and *Baseball References*. We collect information from the 1957 season, the year when the qualified number of plate-appearances was regulated. It is the cutoff point to be eligible for the title of leading hitter, the player with the highest batting-average. Stats in each season cover only the regular season, not Spring-Training or postseason games. The full-sample size is $N = 54469$.

Note that we then restrict with the subsample including the players who appear to the plate in MLB games enough to be tested because those with a small number of plate-appearances are likely to be evaluated by other elements than their batting skills . . . pitching or fielding, performance at the minor leagues, or those who injured at the season. Especially the batting-average and on-base percentage of players with few plate-appearances are not be regarded as reliable. Pope and Simonsohn (2011) set the cutoff at 200 at-bats, but alternatively, we use 200 plate-appearances as the required number to be considered in our analysis. This is because at-bat does not count the number of base-on-balls or sacrifice bunts in the denominator, even though they surely appear to the plate and made something to their teams. On the other hand, plate-appearance stands for the number of chances of batting their coaches gave him. Restricting the sample reduces the sample size to $N = 18143$.

The dataset includes the players' plate-appearance, batting-average, on-base percentage, homerun, stolen-base, runs-batted-in, base-hit, all of which are the main stats of interest in this paper. Besides, for the regression analysis, we obtain additional stats: batting, fielding, and BaseRun, the estimated contribution to the team expressed in the runs they produced, and WPA, or winning percentage added. Further explanations used in our analysis are provided in the Appendix.

We then describe the characteristics of the stats. Baseball batting stats are roughly divided into two types. The first is the one that simply indicates the number of certain plays, and the second is the one calculated using these numbers, which indicates the rate or the expected number of the plays. Here we call the former "count stats," and the latter "rate stats." count stats are for example "base-hit" or "home run," or "stolen-bases." count stats are irreversible and so they are monotonically increasing in their appearances. That is, once the players reached a certain number of count stats such as 20 home runs, it does not matter how their performance goes afterward. On the other hand, rate stats can fluctuate: even once they reached their internal goals, it may fall if their performance becomes worse. Batting-average and on-base percentage are categorized to this type.

Second, we mention the salary data. Salary data are obtained from *USA TODAY* and *Baseball References*. We collected annual salary data of the position players who registered into MLB

Table 1: Summary Statistics for Sample A

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
PA	18,143	456.477	152.836	200	320	591	778
AVG	18,143	.264	.032	.135	.242	.285	.394
OBP	18,143	.331	.039	.174	.305	.356	.609
HR	18,143	11.811	9.747	0	4	17	73
RBI	18,143	51.882	26.912	4	31	69	165
SB	18,143	7.846	10.869	0	1	10	130
H	18,143	108.941	42.933	29	72	143	262
Age	18,143	28.506	4.042	18	25	31	46

Table 2: Summary Statistics for Sample B

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Age	8,915	28.714	3.901	19	26	31	46
PA	8,915	471.946	150.890	200	342	605.5	778
AVG	8,915	.268	.031	.146	.248	.289	.394
OBP	8,915	.337	.038	.174	.311	.360	.609
HR	8,915	13.446	10.213	0	6	19	73
RBI	8,915	56.339	27.621	5	35	74	165
SB	8,915	8.534	10.851	0	1	11	109
H	8,915	114.232	42.481	30	78	148	262
+WPA	8,915	8.715	3.471	2.030	5.820	11.430	19.160
-WPA	8,915	−8.270	2.610	−15.050	−10.420	−6.060	−2.740
Bat	8,915	3.257	16.139	−44.200	−7.300	11.100	116.800
Fld	8,870	.304	7.482	−36.100	−4.000	4.400	37.000
BsR	8,915	.092	2.712	−12.600	−1.200	1.200	14.300
Salary	8,915	3,487,838	4,487,344	62,500	512,750	4,658,334	29,200,000
FA	8,915	.168	.374	0	0	0	1

Roster at the beginning of each season. They also contain information about other player-specific characteristics: age, position (such as catcher, 1st-baseman, left-fielder, designated hitter and so on), the teams they signed, and the possession of the free agency. We merged this dataset with the player's stats in the previous year, because the salary is usually determined based on the performance of the previous season. Because of the lack of disclosed information, this dataset contains data on salary only from 1987. Because we regard the players' performance as reflected in their annual reward in that of the next year, we cannot merge the stats dataset of 2018. So the latest available season is 2017. The aggregated number of the panel is $N = 13226$, and the restriction of 200 plate-appearances reduces the number to $N = 8915$. Then, here we use two datasets: one that contains salary data (we call this Sample B), while the other does not (Sample A). As we explain in the next section, we use two main analyses: manipulation of performance stats (only use play stats) and design of the contracts (needs information about salary). In the former, we use Sample A, and in the latter Sample B, The descriptive statistics of each sample are described in Table 1 and Table 2, respectively.

4 Empirical Methodology

4.1 Test for Bunching

First, we test if we can observe any behavior that seems to be related to reference-point dependent preference. As we explained in the previous sections, manipulation is verified by the observation of bunching or an excess mass around the possible reference point. To specify the excess mass, we apply the method of regression discontinuity design (RDD).

RDD is a way to measure the effect of a treatment, such that whether the treatment is assigned or not depends on the threshold of a certain variable (called "running variable"). Then, comparing the samples just above and just below the threshold is sufficient examination of the treatment, since they are in almost the same states except for the existence of the treatment.

However, there is an important assumption for this specification is regarded to be valid (Lee and Lemieux, 2010): continuity of the running variable around the threshold. In other words, individuals must not be able to manipulate the running variable to cross over the cutpoint. If there exists manipulation, then a selection bias problem arises, that is, those who try to be assigned the treatment can adjust their running variable. Therefore, there are some empirical ways to test the manipulation of a variable, which is the very method I apply in our analysis. One of the frequently applied methods for this specification is McCrary(2008)'s local linear density estimation. We avail of this to our specification of bunching.

This test of the manipulation at the cutoff point c proceeds in two steps. First, let $d(.)$ denote the density function of the variable x to be tested (here we use the performance stats). Then we undersmooth the observed density: determine the binsize b of x, and obtain the histogram. And finally, we conduct local linear approximation to both just above and below the cutoff point. Note that the optimal bandwidth is to be selected by the fourth-order polynomial approximation. Then, we estimate the frequency at the cutoff point, $\hat{d}(r)$, by fitting the estimated density function from both below and above the cutoff, $\hat{d}^{+}$ and $\hat{d}^{-}$, respectively. Finally, we make the difference between $\ln \hat{d}^{+}$ and $\ln \hat{d}^{-}$ to calculate the statistics θ. With θ and its estimator of standard deviation, t-tests can be conducted to specifying manipulation.

4.2 Monetary Incentives

Then, we examine the existence of the monetary incentive. For this procedure we use the specification of the two functional features mentioned in Section 4.1: notch and kink at certain cutoff points.

First, we specify the salary scheme $f(.)$ without a notch at the cutoff. Then the entire salary scheme is expressed as follows:

$$w_{it} = \beta_0 + f(\text{PERF}_{it}) + \beta_2 \text{ABOVE}_{it}$$

If $f(.)$ is linear in the performance stats PERF_{it}, then

$$w_{it} = \beta_0 + \beta_1 \text{PERF}_{it} + \beta_2 \text{ABOVE}_{it}$$

For each player i in the season t, w_{it} is logarithm of their average annual salary in next season $t+1$. PERF_{it} is the value of performance stats (batting-average, on-base percentage,...). ABOVE_{it} is an indicator of the achievement of their goals for the stats. If the estimated value of β_2 is supported to be positive and significant, then we consider there is an additional bonus paid to the player that reached a certain target value of the performance stats, which leads them to bunch.

As the second possibility, the kink at the reference point is specified as follows.

$$w_{it} = \beta_0 + \beta_1 \text{PERF}_{it} + \beta_2 \text{ABOVE}_{it} + \beta_3 \text{PERF}_{it} \times \text{ABOVE}_{it}$$

The first three terms on the right-hand side are the same as the specification of a notch, the linear salary scheme in the performance stats. In addition, we add the interaction term of PERF_{it} and ABOVE_{it}. This term stands for the return to the performance stats for the players that achieve the cutoff point. In words, players with .320 of batting-average, for example, are evaluated by β_1 until .300, and then $\beta_1 + \beta_3$ for the rest .020 (it is the case of that bunching at .300 is argued). If the estimated value of β_3 is negative and significant, then there exists kink in the reward function, which again supports the bunching at the cutoff point.

Then, we describe the empirical specification. In our paper, we specify the local polynomial functions described above by the regression analysis, for each performance stats (batting-average, for example) and cutoff point (.300). We employ the methodology of regression discontinuity design (RDD[1]). We restrict the sample to those who are around the cutoff of each stats (for example .025 around .300 for batting-average). The bandwidths in each analysis are selected according to the optimization of Imbens and Kalyanaraman (2009).

We also use linear regression with the interaction term of PERF_{it} and ABOVE_{it} to see the kink at the reference point. In each analysis, we restrict the sample to the players with their stats around the cutoff. The window sizes are selected ad hoc, so we tried various cutoffs to check the robustness.

$$w_{it} = \beta_0 + \beta_1 \text{PERF}_{it} + \beta_2 \text{ABOVE}_{it} + \beta_3 \text{PERF}_{it} \times \text{ABOVE}_{it}$$

[1] We have to note that there is a problem to apply RDD in this case: The running variable is manipulated. In Section 5., we come back to this topic.

Furthermore, to check the robustness of the results, we include the terms about the player-specific characteristics as well. Thus, the model to be estimated is:

$$w_{it} = \beta_0 + \beta_1 \mathrm{PERF}_{it} + \beta_2 \mathrm{ABOVE}_{it} + \mathbf{bZ}_{it}$$
$$w_{it} = \beta_0 + \beta_1 \mathrm{PERF}_{it} + \beta_2 \mathrm{ABOVE}_{it} + \beta_3 \mathrm{PERF}_{it} \times \mathrm{ABOVE}_{it} + \mathbf{bZ}_{it}$$

Team and individual fixed effect models are also included.

One important problem of this method is the lack of consideration of the possibility of endogeneity about our results. We discuss this in Section 5.

5 Result

5.1 Excess Mass Around The Reference Point

In this section, we present our main results of the analysis. First, we show the results that verify bunching. Table 3 includes the summary of McCrary (2008)'s manipulation tests about the performance stats. Consistent with Pope and Simonsohn (2011), there exists excess mass around .300, and besides .250 of batting-average. Manipulation was also observed in some of the other round numbers of other stats: .350 of on-base percentage, 20 of home runs, 100 of runs-batted-in, 30 and 40 of stolen-bases, and 200 of base-hits.

For precise estimation of bunching, we set the binsizes of undersmoothing artificially: .001 for batting-average and on-base percentage, 1 for homerun (HR), stolen-base (SB), plate-appearance (PA), and base-hit (H), and 4 for runs-batted-in (RBI). Batting-average (AVG) and on-base percentage (OBP) are usually shown by three decimal digits, rounding the fourth decimal digit, so strictly batters with .2995 of batting-average are taken as .300. As we mentioned in Section 3, home run, stolen-base, plate-appearance, and base-hit stand for the number of plays in interest, and they, therefore, take an integer and earn one for each plate-appearance. Runs-batted-in is also an integer-stats, but it can get at most 4 at one plate-appearance, so we set it 4. To confirm the robustness of our results, we repeated this test with various binsize, but we yielded essentially the same results. Bandwidths are optimized by calculation, following McCrary (2007).

For .300 of batting-average, when we increased the sample size compared with Pope and Simonsohn (2011), we obtained similar results: Players put effort to manipulate their batting-average. A difference between the estimated frequency according to the approximation below .300 and that of above .300 was significant at the .1% ($z = 7.442$) level.

In addition, bunching occurs in .250 ($z = 5.061$, $p < .1\%$) as well. It was not reported in Pope and Simonsohn (2011): there was no bunching observed in any other round numbers of batting-average, so bunching may have occurred before 1973 (Pope and Simonsohn (2011)

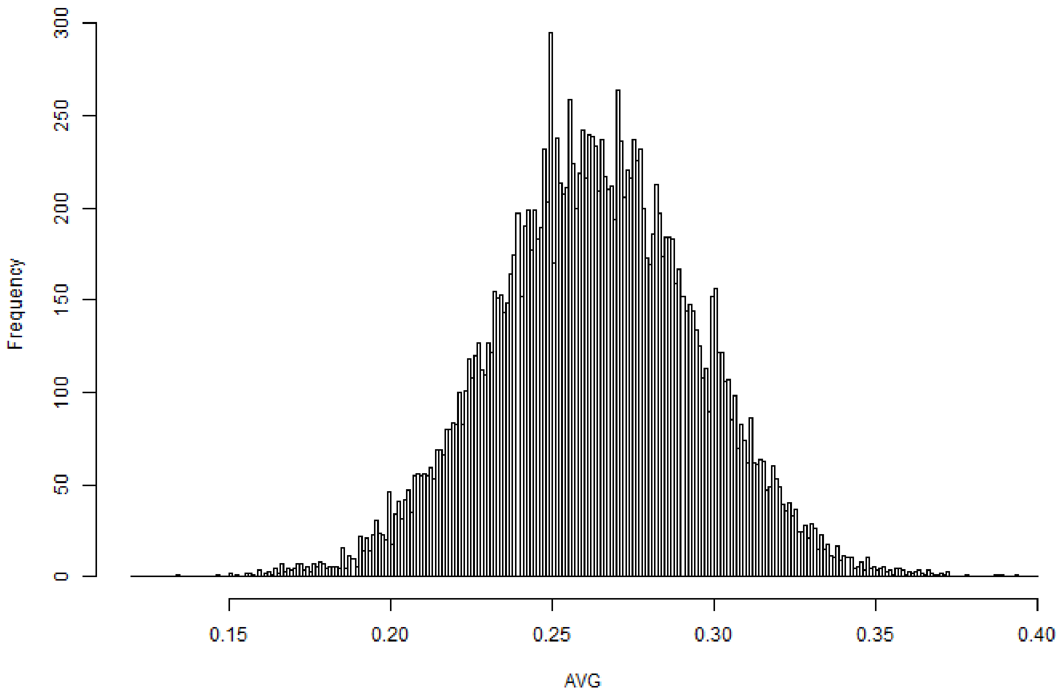

Figure 4: Histgram of Batting-Average

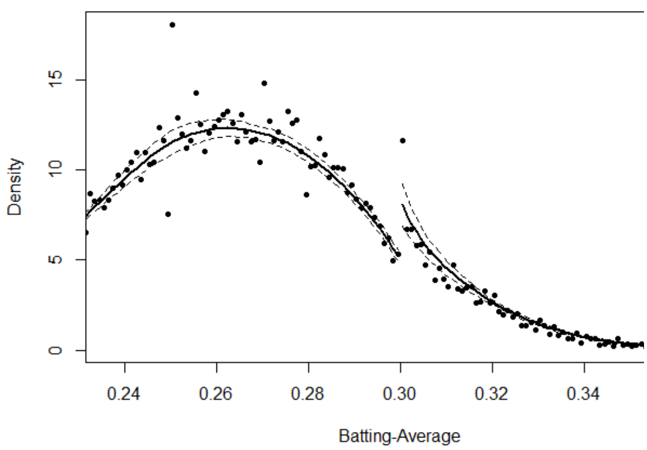

Figure 5: Discontinuity at .300 of AVG

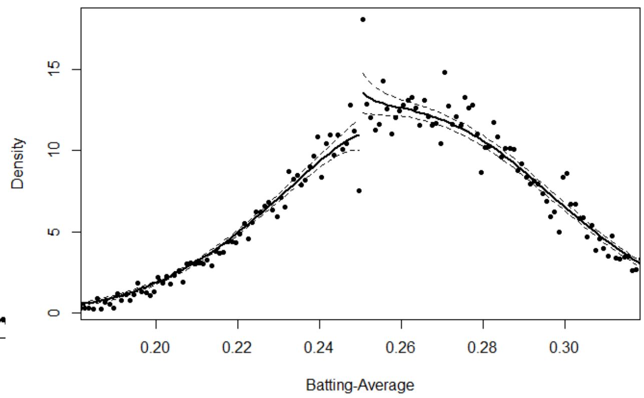

Figure 6: Discontinuity at .250 of AVG

Table 3: Test for Manipulation, leastPA $= 200$

stats	type	cutpoint	binsize	bandwidth	θ	z
AVG	rate	.300	.001	.019	.499	7.442***
					(.067)	
		.250	.001	.024	.212	5.061***
					(.042)	
OBP	rate	.350	.001	.024	.139	2.854**
					(.049)	
HR	count	20	1	5.309	.259	3.465***
					(.075)	
RBI	count	100	4	15.423	.311	3.295***
					(.094)	
SB	count	30	1	10.000	.529	4.274***
					(.124)	
		40	1	11.505	.481	2.764**
					(.174)	
PA	count	500	1	.003	.160	2.515*
					(.063)	
H	count	200	1	18.922	.453	2.547 *
					(.178)	

Note ***: $p < 0.1\%$, **: $p < 1\%$, *: $p < 5\%$.
Bandwidth is optimized following the method of McCrary(2008).

includes dataset after 1974). We specifically analyzed this in Section 6.4. Replicating analysis with larger binsizes: .002 and .005 also yielded similar results.

On-base percentage, on the other hand, showed a similar tendency in .350, at the significance level was the 5% level ($z = 2.854$). Pope and Simonsohn (2011) reported that when their batting-average was just below .300 at the last plate-appearance of the season, they tried to avoid getting walks to get a base-hit. This implies that players regard the batting-average as more important than the on-base percentage. Our results, however, suggest that to some extent, players set goals about their on-base percentages as well.

Bunching also occurs in count stats. They showed similar results. As well as batting-average and on-base percentages, bunching is observed only in some round-numbers, not in all of them. For Homerun, bunching occurs only in 20 ($z = 3.465, p < 0.1\%$): there may be diminishing sensitivity: 20 is located on above the 75 percentiles of the whole Sample A. We observe bunching also at 30 ($z = 4.274, p < 0.1\%$) and 40 ($z = 2.764, p < 1\%$) of stolen-base and 200 in base-hit ($z = 2.547, p < 5\%$). Stealing-bases require skills that are talented to a limited number of the players but succeed with the probability of 60% to almost 100%, so those who are evaluated by their number of stolen-bases can manipulate them. 30 and 40 of stolen-bases

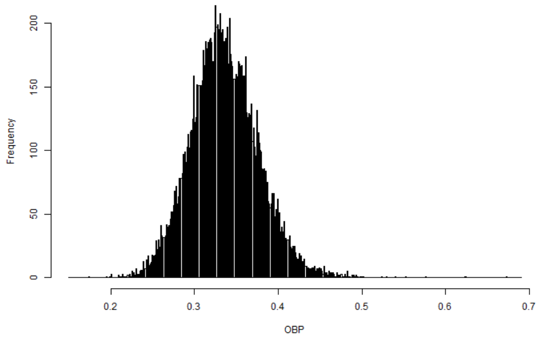

Figure 7: Histgram of On-Base Percentage

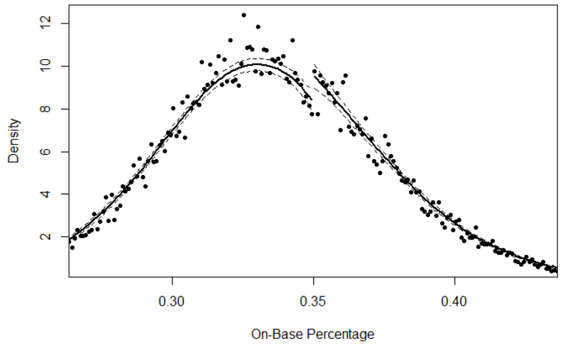

Figure 8: Discontinuity at .350 of OBP

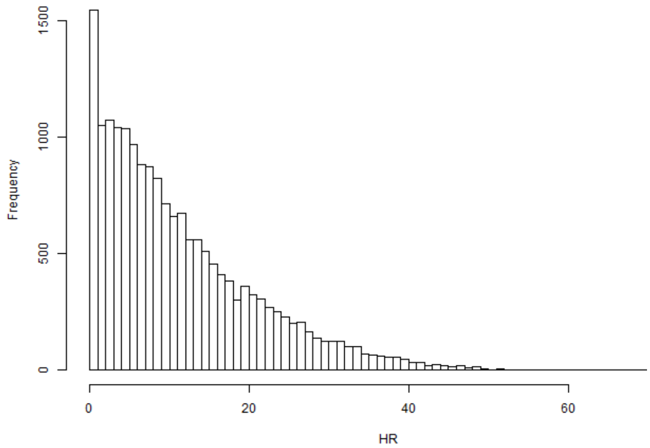

Figure 9: Histgram of Homerun

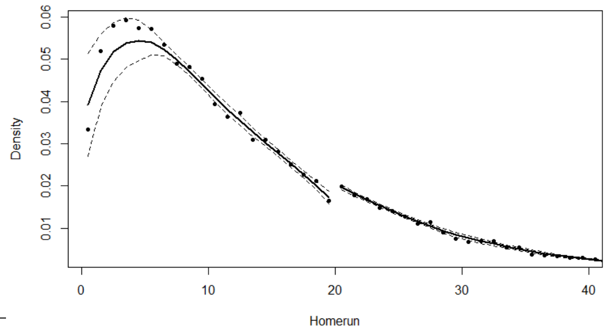

Figure 10: Discontinuity at 20 of HR

are also far above the 75 percentiles of all the players.

Base-hit was also manipulated, but the confidence level of the discontinuity was lower than that of batting-average (z = 2.547, $p < 5\%$). Base-hit is a stats close to batting-average because both stats increase by getting base-hits. We obtain this result because the number of base-hit is not regarded as important as batting-average (Most TV live on baseball display a player's batting-average, not the number of base-hits.) Alternatively, we can consider that this is the difference between the count stats and the rate ones. If a player reaches .300 of batting-average, for example, then he can keep it by not attending the rest of the plate-appearances or games. Pope and Simonsohn (2011) pointed out that players with just above .300 were more likely to be replaced with their last scheduled plate-appearance. On the other hand, if he gets the 200th base-hit, then he does not have to care about the number of base-hits and attend the games to get better performance.

Surprisingly, such manipulation occurs in runs-batted-in (z=3.295, p<0.1%), too. Compared

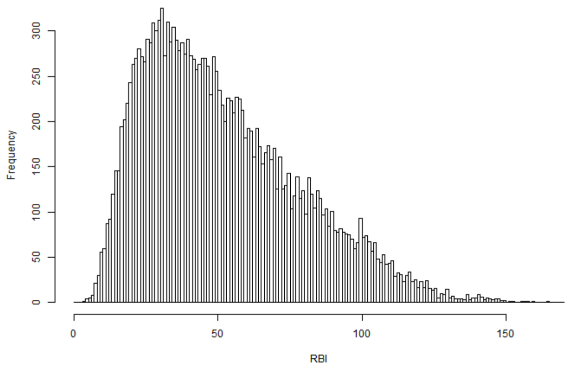

Figure 11: Histgram of Runs-Batted-In

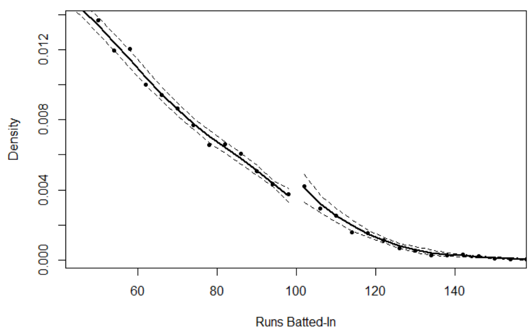

Figure 12: Discontinuity at 100 of RBI

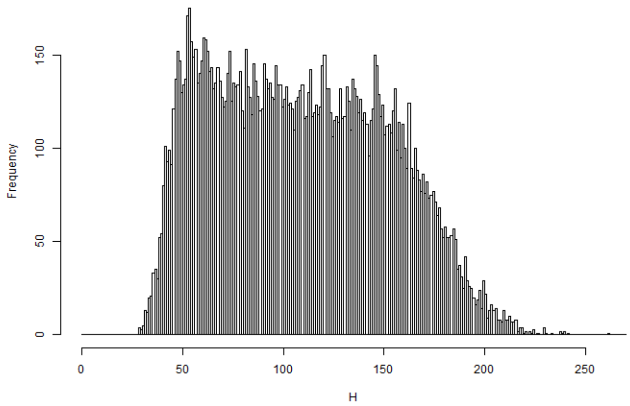

Figure 13: Histgram of Base-Hit

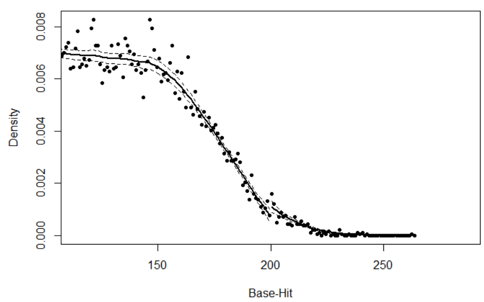

Figure 14: Discontinuity at 200 of Base-Hit

to other stats, runs-batted-in is harder to be manipulated because the number of that depends on the performance of his teammates, and the number they can earn at a single plate-appearance varies from 1 to 4. As in Table 3, there occurs the evidence in plate-appearances. However, it may not be convincing because the optimized binwidth was smaller than one, even though the number of PAs takes only integers. Setting binwidth larger than 1, then the result turned out to be insignificant, so we regard it as not supportive results for bunching. (In fact, there certainly exist monetary incentives for plate appearances. We come back to here in Section 6.1.)

Summarizing the results, there exists bunching at some round-numbers of the batting stats and they are the possible reference point of the players. In the case of marathon runners, Allen et al. (2016), there occurred bunching in every round number of the goal time, although the size of discontinuity decreased. That is, it should be considered that the reference points are not determined only by round numbers as argued by Pope and Simonsohn (2011). That is, the nature of the reference points are likely to be close to "par" in Pope and Schweizer (2011), or the well-known standards that are related to the image of "skilled players," rather than round

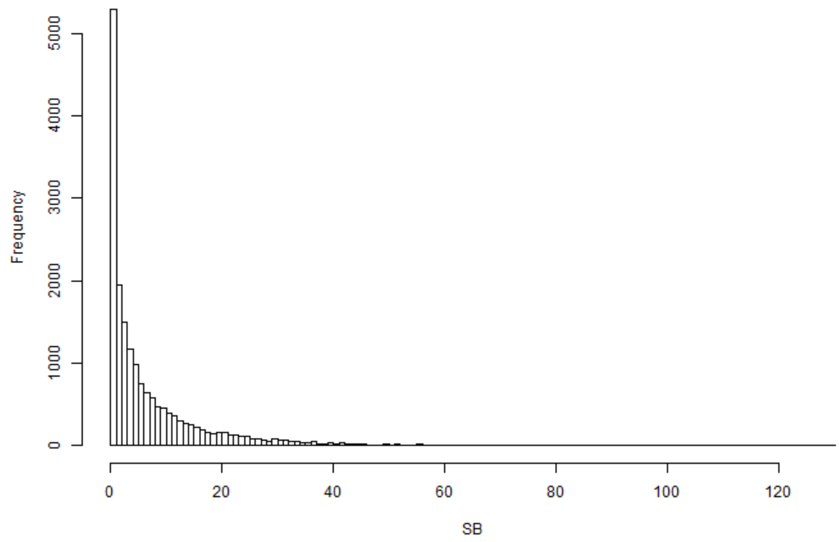

Figure 15: Histgram of Stolen-Base

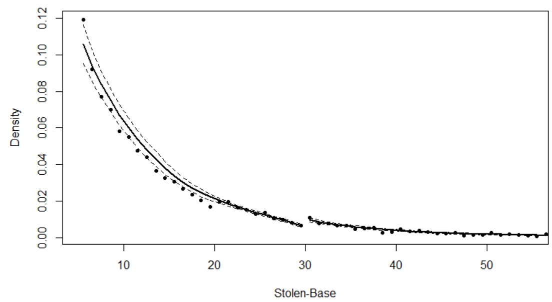

Figure 16: Discontinuity at .300 of AVG

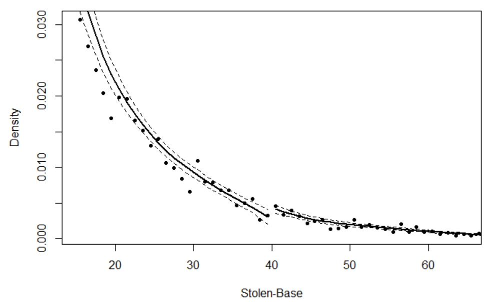

Figure 17: Discontinuity at .250 of AVG

numbers; or well, these numbers are monetarily incentivized goals by the team managers. So next, we examine whether there is any monetary bonus in their contract.

5.2 Existence of Monetary Incentive

In Section 5.1, we confirmed the discontinuities at the cutoff points of the representative stats. Then in this section, we show whether they target these goals based on their reference dependence or their design of the contracts with a monetary incentive.

Table 4 describes the results of local linear regressions on the logarithm of their salary next year, with the cutoffs of each possible reference point. Column "Other Control" indicates "Yes" if the model includes covariates (Defense, BaseRun, player's age, WPA and dummy for possession of the right of free agency) and season dummies, but "No" otherwise. "bw type" indicates the bandwidths used in the model: “Optimal” includes the sample that is in the optimal bandwidth calculated by Imbens and Kalyanaraman (2009), while “Half-BW” and “Double-BW” are using a half and a twice of the Optimal bandwidth, respectively. The local average treatment effect of achieving the goals are described in the sixth column.

As a whole, no evidence supports the existence of monetary incentive to make an effort for their observed goals. There is no essential difference between the count stats and rate stats. That is, the reward function $f(.)$ does not discontinuously jump at each cutoff point.

The other possibility that encourages players to manipulate stats is the kink of the monetary incentives. Table 5 to 12 show results of local linear regression including the interaction terms of the dummy variable and those of the corresponding stats, or X_{it} and ABOVE$_{it}$. "PA" is the number of plate-appearance. The second row is the performance stats, and the third one is ABOVE$_{it}$ for the corresponding possible reference points. The fourth one is the interaction terms.

Results again rejected the hypothesis of the existence of monetary incentives for bunching. There is no evidence of neither kink or jump of their reward functions. Here we consider each stats respectively.

First, for batting-average, results rejected the hypothesis of the existence of any additional monetary bonus for achieving either .300 or .250. Although estimated jumps at each cutoff point were positive, their standard errors are large and so the difference was insignificant. The same results were obtained in the model with interaction terms (Tables 5 and 6): the estimated coefficients of the dummies for achieving their internal goals and the interaction term with the batting-average are all insignificant. That is, the monetary reward does not discontinuously jump or kink at .300 of .250. These findings are consistent with the hypothesis that preferences of the players are reference-dependent about their batting-average and excess mass is caused by reference-dependent preferences, not by the monetary incentives.

On-base percentage seems similar. The estimated jump is negative, although statistically insignificant. Table 7 shows that negative kink is not observed. On-base percentage is considered as relatively more important stats: it is a closer correlation with the winning-average of the team than batting-average. Thus, it can be the case that team managers attach a higher value to on-base percentage than to batting-average and think of paying players with a higher number more. However, our results are against this hypothesis.

For the count stats, observed results are similar: 20 of homerun, 100 of runs-batted-in, and 200 base-hits work as the cutoff points of bunching, but not as the marginal points for the players' salary. Homeruns produce at least one score to the team and take one of the most “attractive” aspects of baseball, so there may exist an additional positive effect for the team: it may bring a lot of audiences, which profits them by stadium fees. Nevertheless, the discontinuous scheme of the salary was not observed. Regressions with the interaction term do not report kink at each cutoff point.

Stolen-base, however, shows different results. Some of the results from Table 4 showed a significant discontinuity of about 30 stolen-bases. According to Tables 11 and 12, some of the regressions (column (1) for 30 SB and (1) and (6) for 40 SB) support the jump and negative kink at the cutoff points. That is, not all the results are consistent with the analyses of other stats. However, we do not regard them as sufficient support. First, in Table 4, controlling other player-specific characteristics, their significance level drastically goes down. In Table 11, the estimated values are a mixture of significant ones and insignificant ones, and they fluctuate from .816 to 13.567. And finally, for 40 stolen-bases, the results of estimation showed a negative jump, inconsistent with that of 30, even though these results argue the same stats. Thus, we conclude these results cannot support the hypothesis of no jump or kink of the salary contracts $f(.)$, but either vice versa. Especially at this point, we require further analysis.

One possible alternative interpretation is that there exit the players that sign the contracts that include plural-year service. Such a player receives a fixed salary regardless of their single-year performance. Thus, we restrict the sample to those who have the right of free agency, which enables them to negotiate with any MLB or other professional baseball teams. These players just have finished their contract in the season, and so cannot play for any MLB teams without signing a new contract, which always reflects his performance of the previous year. In the analysis above, we consider the possession of the right of free agency by adding the dummy variable that indicates whether he holds the right or not.

Table 13 shows the results of local-linear regression, using the restricted sample of free-agent players. This is consistent with the main results: there does not exist evidence that supports the additional reward at each cutoff point. Analysis of stolen-bases was not conducted because of a lack of players around there. Table 14 shows the results about .300 of batting-average, also for the free-agent players. The obtained implication is the same as the first analysis: there is no kink and notch in the reward function.

5.3 Endogeneity Problem

One important factor to explain the results is the possible endogeneity problem. This is caused by the methodology we applied in the analysis above.

As we mentioned above, we partly used the method of RDD to specify monetary incentives. The required condition for this analysis is that the running variable is not manipulated by the individuals. RDD design assumes that individuals just above and just below the cutoff are almost the same except for that one is reaching the cutoff. If some players know the jump or kink in $f(.)$ at the .300 of batting-average, they try to keep their batting-averages above thereby avoiding

Table 4: Linear Regression on Monetary Incentives

stats,cutpoint	Other Control	bw type	bandwidth	Observations	Estimate	Std. Error	z
AVG, .300	No	Optimal	.084	8514	.047	.061	.773
		Half-BW	.042	5599	.088	.075	1.174
		Double-BW	.170	8915	.067	.056	1.184
	Yes	Optimal	.045	5930	.034	.056	.615
		Half-BW	.023	3005	.061	.077	.788
		Double-BW	.090	8605	.016	.045	.354
AVG, .250	No	Optimal	.036	6110	.019	.068	.286
		Half-BW	.018	3496	.015	.092	.161
		Double-BW	.072	8539	.034	.054	.636
	Yes	Optimal	.048	7271	.070	.052	1.340
		Half-BW	.024	4402	.066	.069	.953
		Double-BW	.096	8810	.075	.044	1.713
HR, 20	No	Optimal	3.32	1315	.071	.175	.406
		Half-BW	1.66	562	.073	.127	.576
		Double-BW	6.64	2582	-.004	.109	-.034
	Yes	Optimal	3.30	1307	-.002	.141	-.015
		Half-BW	1.65	560	.030	.102	.299
		Double-BW	6.61	2558	-.032	.088	-.364
OBP, .350	No	Optimal	.044	6440	-.038	.065	-.592
		Half-BW	.021	3542	-.076	.089	-.849
		Double-BW	.087	8656	-.029	.051	-.570
	Yes	Optimal	.045	6525	-.013	.049	-.272
		Half-BW	.022	3673	-.055	.069	-.807
		Double-BW	.089	8637	.004	.039	.107
RBI, 100	No	Optimal	4.08	393	.072	.289	.250
		Half-BW	2.04	228	.282	.400	.707
		Double-BW	8.16	714	.008	.185	.043
	Yes	Optimal	4.04	390	.018	.209	.086
		Half-BW	2.02	227	-.042	.324	.130
		Double-BW	8.07	708	.056	.127	.435
H, 200	No	Optimal	3.173	75	-.786	.396	-1.985*
		Half-BW	1.587	35	.386	.271	-1.421
		Double-BW	6.347	137	-.061	.309	-.199
	Yes	Optimal	3.175	75	-.420	1.042	-.403
		Half-BW	1.587	35	-4.779	.576	-8.288**
		Double-BW	6.349	137	-.109	.413	-.265
SB, 30	No	Optimal	3.39	282	.962	.372	2.585**
		Half-BW	1.70	134	.920	.263	3.492***
		Double-BW	8.16	714	.008	.185	2.941**
	Yes	Optimal	3.40	282	.379	.297	1.271
		Half-BW	1.70	134	.290	.249	1.163
		Double-BW	6.79	533	.408	.180	2.260*
SB, 40	No	Optimal	3.16	134	-1.276	.453	-2.818**
		Half-BW	1.58	56	-.736	.383	-1.924
		Double-BW	6.32	245	-.712	.313	-2.274*
	Yes	Optimal	3.16	134	-.346	.396	-.875
		Half-BW	1.58	56	-.313	.429	-.730
		Double-BW	6.33	245	-.115	.244	-.472

Note: ***: $p < 0.1\%$, **: $p < 1\%$, *: $p < 5\%$.
Bandwidth is optimized following the method of Imbens-Kalyanaraman (2009).

Table 5: Local Linear Regression on Monetary Incentives

	(1)	(2)	(3)	(4)
(Intercept)	11.269 ***	-3.374 ***	-2.122 ***	
	(0.386)	(0.631)	(0.635)	
PA	0.004 ***	0.004 ***	0.004 ***	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)
AVG	3.794 **	3.677 **	0.172	0.387
	(1.418)	(1.209)	(1.219)	(1.010)
AVG_300	-1.081	-0.949	-0.394	-0.708
	(0.920)	(0.815)	(0.799)	(0.647)
AVG * AVG_300	3.851	3.312	1.414	2.477
	(3.009)	(2.663)	(2.611)	(2.110)
Age		0.885 ***	0.874 ***	0.901 ***
		(0.037)	(0.036)	(0.031)
Age^2		-0.013 ***	-0.013 ***	-0.013 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	5960	5960	5930	5930
R squared	0.237	0.431	0.453	0.642
F statistic	519.892	855.842	520.717	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 6: Local Linear Regression on Monetary Incentives, .250 of Batting-Average

	(1)	(2)	(3)	(4)
(Intercept)	12.923 ***	-2.363 ***	-1.746 **	
	(0.465)	(0.639)	(0.644)	
PA	0.004 ***	0.004 ***	0.003 ***	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)
AVG	-2.561	-3.677 *	-5.514 **	-3.425 *
	(2.004)	(1.736)	(1.719)	(1.344)
AVG_250	-1.404 *	-1.753 ***	-1.329 **	-0.904 *
	(0.566)	(0.487)	(0.483)	(0.385)
AVG * AVG_250	5.698 *	7.180 ***	5.480 **	3.783 *
	(2.315)	(1.993)	(1.974)	(1.568)
Age		0.942 ***	0.937 ***	0.952 ***
		(0.034)	(0.034)	(0.028)
Age^2		-0.014 ***	-0.013 ***	-0.014 ***
		(0.001)	(0.001)	(0.000)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	7307	7307	7271	7271
R squared	0.213	0.424	0.444	0.643
F statistic	544.842	1040.701	621.448	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2.01e+03).

Table 7: Local Linear Regression on Monetary Incentives, .350 of On-Base Percentage

	(1)	(2)	(3)	(4)
(Intercept)	10.861 ***	-4.325 ***	-3.212 ***	
	(0.457)	(0.639)	(0.644)	
PA	0.004 ***	0.004 ***	0.004 ***	0.001 ***
	(0.000)	(0.000)	(0.000)	(0.000)
OBP	4.720 ***	3.504 **	0.488	2.218 *
	(1.407)	(1.227)	(1.249)	(1.026)
OBP_350	-0.824	-0.112	0.288	-0.275
	(0.796)	(0.685)	(0.680)	(0.558)
OBP * OBP_350	2.274	0.295	-0.867	0.745
	(2.254)	(1.943)	(1.930)	(1.581)
Age		0.944 ***	0.938 ***	0.962 ***
		(0.035)	(0.035)	(0.029)
Age^2		-0.014 ***	-0.013 ***	-0.014 ***
		(0.001)	(0.001)	(0.000)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	6656	6656	6620	6620
R squared	0.220	0.419	0.435	0.632
F statistic	528.545	927.252	540.942	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

the scheduled plate-appearances, for example. If the team managers that offer the contract can detect those who manipulate stats, the contracts seem worse than the observed number of each stats. That is, the effect of achieving the goals might be underestimated.

Table 8: Local Linear Regression on Monetary Incentives, 20 Homeruns

	(1)	(2)	(3)	(4)
(Intercept)	13.540 ***	-5.255 ***	-4.906 **	
	(1.094)	(1.570)	(1.545)	
PA	0.003 ***	0.003 ***	0.005 ***	0.003 ***
	(0.000)	(0.000)	(0.001)	(0.001)
HR	-0.024	0.015	0.018	0.032
	(0.061)	(0.052)	(0.052)	(0.049)
HR_20	-1.422	-0.918	-0.863	-0.367
	(1.360)	(1.157)	(1.143)	(1.059)
HR * HR_20	0.074	0.043	0.040	0.013
	(0.072)	(0.061)	(0.060)	(0.056)
Age		1.110 ***	1.091 ***	1.113 ***
		(0.081)	(0.079)	(0.070)
Age^2		-0.016 ***	-0.016 ***	-0.016 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	1315	1315	1307	1307
R squared	0.075	0.356	0.386	0.570
F statistic	26.693	125.158	80.250	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 9: Local Linear Regression on Monetary Incentives, 100 Runs-Batted-in

	(1)	(2)	(3)	(4)
(Intercept)	14.542 ***	-5.262	-5.563	
	(3.858)	(3.786)	(3.762)	
PA	0.002 *	0.003 ***	0.005 ***	0.003 *
	(0.001)	(0.001)	(0.001)	(0.001)
RBI	-0.005	-0.007	-0.006	-0.005
	(0.040)	(0.034)	(0.034)	(0.028)
RBI_100	-9.535	-6.213	-5.312	-0.929
	(6.248)	(5.703)	(5.720)	(5.132)
RBI * RBI_100	0.096	0.063	0.054	0.010
	(0.063)	(0.057)	(0.057)	(0.051)
Age		1.217 ***	1.253 ***	1.122 ***
		(0.135)	(0.146)	(0.146)
Age^2		-0.019 ***	-0.019 ***	-0.017 ***
		(0.002)	(0.002)	(0.002)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	476	476	470	470
R squared	0.025	0.286	0.337	0.655
F statistic	2.914	27.138	22.314	
P value	0.021	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 10: Local Linear Regression on Monetary Incentives, 200 Base-Hits

	(1)	(2)	(3)	(4)
(Intercept)	-10.720	-20.591	-15.364	
	(41.586)	(45.800)	(52.998)	
PA	0.006	0.005	0.006	-0.004
	(0.004)	(0.005)	(0.006)	(26.412)
H	0.111	0.083	0.046	0.584
	(0.211)	(0.227)	(0.257)	(3066.727)
H_200	47.353	35.346	20.664	94.406
	(48.323)	(53.189)	(55.472)	(548282.725)
H * H_200	-0.237	-0.177	-0.103	-0.479
	(0.243)	(0.267)	(0.279)	(2776.812)
Age		1.070	1.121	2.640
		(0.618)	(0.717)	(5465.018)
Age^2		-0.018	-0.018	-0.043
		(0.011)	(0.013)	(93.699)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	75	75	75	75
R squared	0.042	0.119	0.207	0.936
F statistic	0.622	1.122	1.268	
P value	0.648	0.359	0.264	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 11: Local Linear Regression on Monetary Incentives, 30 Stolen-Bases

	(1)	(2)	(3)	(4)
(Intercept)	12.892 ***	-4.331	-4.535	
	(2.013)	(2.910)	(2.678)	
PA	0.005 ***	0.005 ***	0.003 *	0.002
	(0.001)	(0.000)	(0.001)	(0.001)
SB	-0.045	-0.085	-0.056	-0.065
	(0.074)	(0.065)	(0.062)	(0.069)
SB_30	6.364 *	1.687	2.209	0.632
	(2.725)	(2.444)	(2.243)	(2.531)
SB * SB_30	-0.191 *	-0.038	-0.060	-0.011
	(0.094)	(0.084)	(0.078)	(0.088)
Age		1.154 ***	1.104 ***	1.132 ***
		(0.164)	(0.156)	(0.170)
Age^2		-0.017 ***	-0.017 ***	-0.017 ***
		(0.003)	(0.003)	(0.003)
Other Controls				X
Fixed effects			Season	Season, Team, Position
#observations	361	361	361	361
R squared	0.232	0.456	0.547	0.751
F statistic	27.843	58.663	52.213	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

5.3.1 Donut-RDD

To deal with this possible endogeneity, we repeat the analysis above without the players in the bandwidth of .005 of batting-average. .005 is the obtained value by a base-hit for the player with 200 at-bats. Table 15 shows the results.

We report the same result: the coefficients of the dummy variable and the interaction term

Table 12: Local Linear Regression on Monetary Incentives, 40 Stolen-Bases

	(1)	(2)	(3)	(4)
(Intercept)	6.138	-4.880	0.455	
	(3.304)	(4.399)	(4.444)	
PA	0.004 ***	0.004 ***	0.003	0.005
	(0.001)	(0.001)	(0.002)	(0.004)
SB	0.167	0.120	0.089	-0.007
	(0.087)	(0.084)	(0.081)	(0.143)
SB_40	14.200 **	12.501 *	9.582 *	7.506
	(5.387)	(4.878)	(4.666)	(7.472)
SB * SB_40	-0.370 **	-0.321 **	-0.252 *	-0.186
	(0.135)	(0.123)	(0.118)	(0.193)
Age		0.772 ***	0.466 *	0.935 **
		(0.210)	(0.227)	(0.301)
Age^2		-0.011 **	-0.006	-0.014 **
		(0.003)	(0.004)	(0.005)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	167	167	167	167
R squared	0.215	0.365	0.520	0.781
F statistic	10.901	14.658	13.078	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** p $<$ 0.001; ** p $<$ 0.01; * p $<$ 0.05. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 13: Local Linear Regression on Monetary Incentives for FA Players

stats,cutpoint	Other Control	bw type	bandwidth	Observations	Estimate	Std. Error	z
AVG, .300	No	Optimal	.025	503	-.175	.197	-.888
		Half-BW	.013	252	-.307	.302	-1.016
		Double-BW	.052	1043	-.180	.141	-1.271
	Yes	Optimal	.026	509	-.253	.138	-1.832
		Half-BW	.013	266	-.209	.212	-.986
		Double-BW	.052	1038	.199	.102	-1.938
AVG, .250	No	Optimal	.056	1366	.074	.102	.721
		Half-BW	.028	910	.147	.133	1.099
		Double-BW	.114	1501	.067	.090	.735
	Yes	Optimal	.058	1367	.084	.082	1.020
		Half-BW	.029	923	.149	.107	.398
		Double-BW	.117	1480	.070	.072	.964
HR, 20	No	Optimal	3.48	211	-.302	.300	-1.007
		Half-BW	1.74	96	-.123	.226	-.543
		Double-BW	6.96	387	.045	.203	.224
	Yes	Optimal	3.50	206	-.273	.296	-.924
		Half-BW	1.75	95	-.156	.278	-.560
		Double-BW	7.00	439	-.098	.174	-.565
OBP, .350	No	Optimal	.045	1103	.034	.129	.262
		Half-BW	.023	597	-.106	.172	-.620
		Double-BW	.092	1469	.024	.105	.225
	Yes	Optimal	.043	1044	.021	.107	.196
		Half-BW	.021	566	-.085	.153	-.558
		Double-BW	.086	1435	.016	.084	.194
RBI, 100	No	Optimal	4.90	50	-.100	.559	-.179
		Half-BW	2.45	30	-.095	.949	-.101
		Double-BW	9.80	102	.256	.333	.770
	Yes	Optimal	4.93	49	.195	.433	.449
		Half-BW	2.46	30	-1.360	1.295	-1.050
		Double-BW	9.86	100	.398	.160	2.481*
H, 200	No	Optimal	4.498	107	-.070	.447	-.156
		Half-BW	2.249	61	-.439	.726	-.605
		Double-BW	8.996	218	-.025	.293	-.086
	Yes	Optimal	4.512	106	.649	.355	1.824
		Half-BW	2.256	61	1.084	.963	1.125
		Double-BW	9.024	240	.264	.243	1.087

Note: ***: $p < 0.1\%$, **: $p < 1\%$, *: $p < 5\%$.
Bandwidth is optimized following the method of Imbens-Kalyanaraman (2009).
For stolen-bases, it cannot be optimized because of lack of samples.

Table 14: For FA players: Local Linear Regression on Monetary Incentives

	(1)	(2)	(3)	(4)
(Intercept)	6.803 ***	6.827 *	8.564 ***	
	(1.887)	(3.000)	(2.432)	
PA	0.004 ***	0.004 ***	0.002	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	20.329 **	21.148 **	12.552 *	6.042
	(6.622)	(6.699)	(6.201)	(4.675)
AVG_300	7.053 *	7.289 *	7.661 *	2.372
	(3.266)	(3.267)	(3.100)	(2.296)
AVG * AVG_300	-23.807 *	-24.652 *	-25.787 *	-7.886
	(10.864)	(10.870)	(10.311)	(7.658)
Age		0.009	0.055	0.132
		(0.150)	(0.111)	(0.101)
Age^2		-0.000	-0.001	-0.002
		(0.002)	(0.002)	(0.001)
WAR3			0.090 ***	0.090 ***
			(0.012)	(0.009)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	503	503	492	492
R squared	0.373	0.378	0.522	0.824
F statistic	82.744	56.353	47.385	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 15: Donuts RDD: .300 of Batting-Average

	(1)	(2)	(3)	(4)
(Intercept)	11.640 ***	-3.487 ***	-2.397 ***	
	(0.426)	(0.683)	(0.687)	
PA	0.004 ***	0.004 ***	0.004 ***	0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)
AVG	2.412	2.904 *	-0.464	0.526
	(1.575)	(1.340)	(1.349)	(1.121)
AVG_300	-3.049 *	-2.262 *	-1.597	-1.109
	(1.249)	(1.119)	(1.086)	(0.890)
AVG * AVG_300	10.157 *	7.501 *	5.233	3.721
	(3.997)	(3.574)	(3.474)	(2.841)
Age		0.908 ***	0.904 ***	0.923 ***
		(0.039)	(0.039)	(0.033)
Age^2		-0.013 ***	-0.013 ***	-0.013 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	5259	5259	5232	5232
R squared	0.234	0.430	0.452	0.642
F statistic	451.710	744.160	455.735	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

with the batting-average are statistically insignificant. That is, there is no jump or kink in their reward functions. Analysis of other cutoff points yields the same results.

5.3.2 Stats before Manipulation

Furthermore, we checked robustness by using sample before manipulation.

Table 16: For FA players, donuts RDD: .300 of Batting-Average

	(1)	(2)	(3)	(4)
(Intercept)	7.929 ***	8.320 *		
	(2.338)	(3.494)		
PA	0.004 ***	0.004 ***	-0.001	-0.000
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	16.177	17.519 *	12.830 *	12.779 *
	(8.271)	(8.418)	(5.185)	(5.851)
AVG_300	1.360	1.633	1.099	1.546
	(5.017)	(4.886)	(3.368)	(3.564)
AVG * AVG_300	-5.371	-6.392	-4.712	-6.098
	(16.353)	(15.962)	(10.948)	(11.645)
Age		-0.018	0.114	0.169
		(0.165)	(0.105)	(0.132)
Age^2		-0.000	-0.002	-0.003
		(0.002)	(0.002)	(0.002)
WAR3			0.096 ***	0.089 ***
			(0.011)	(0.012)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	397	397	389	389
R squared	0.372	0.379	0.795	0.823
F statistic	63.013	43.812		
P value	0.000	0.000		

Linear regressions on loggarithm of players' salary next season.
Note: *** p $<$ 0.001; ** p $<$ 0.01; * p $<$ 0.05. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

The objective of the players is to finish their season with their stats above the cutoffs, so they start manipulation just before the end of each season.

Figure 17 shows the histogram of batting-average after 2002[2], at the end of the games of September 28th of each season. The regular seasons of MLB usually end by September, so on 28th, each team remain two or three games. Generally, players manipulate their barring-averages by avoiding the last several plate-appearances at the end of each season: that is, this subsample captures stats before manipulation.

Figure 18 displays the result of McCrary (2008)'s manipulation test.The figure in the left shows the result on September 28th, while the other is that at the end of each season. No bunching is found at .300 ($z = 1.10, p = 27\%$) but at the end of the season, players manipulated their batting-averages ($z = 4.65, p < .1\%$).

Table 17 shows the result of local linear regression using the subsample above. Again, at least, there are no statistically significant effects found for achieving the round-number, .300. Restricting the subsample to those with free agency yields essentially the same results.

5.3.3 Players Unable to Manipulate the Stats

Another point to consider the robustness is that manipulation is usually likely to take place among players on the teams with their standings having been confirmed.

[2] This is because of the limitation of data-availability.

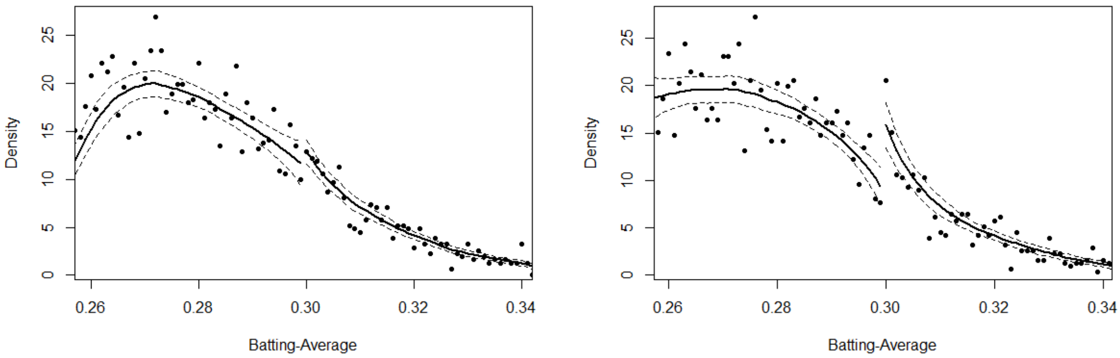

Figure 18: Before-After: Manipulation Test

Table 17: Stats with "before" manipulation: .300 of batting-average

	(1)	(2)	(3)	(4)
(Intercept)	13.037 ***	-2.700	0.058	
	(2.835)	(2.486)	(2.414)	
PA	0.004 ***	0.004 ***	0.003 ***	0.002 *
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	-0.534	-5.334	-12.007	-10.490
	(9.706)	(7.592)	(7.348)	(7.189)
AVG_300	0.717	-0.117	-0.027	-0.054
	(4.424)	(3.539)	(3.445)	(3.297)
AVG * AVG_300	-1.979	0.849	0.514	0.583
	(14.727)	(11.761)	(11.436)	(10.947)
Age		1.038 ***	0.993 ***	1.005 ***
		(0.073)	(0.070)	(0.071)
Age^2		-0.015 ***	-0.014 ***	-0.014 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	1142	1142	1141	1141
R squared	0.193	0.460	0.508	0.598
F statistic	75.194	183.063	127.631	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

The objectives of the MLB teams are to win the qualification for the postseason game, not to help players to achieve their internal goals, such as .300 of batting-averages. If teams contend until the end of the seasons, it is natural that players are not allowed to manipulate their stats by missing the games. Of course, the audience sees if players try to manipulate them, so there is no information asymmetry. In other words, players cannot secretly manipulate them.

Table 18 shows the results when restricting the sample to the players on the teams with close competition, which satisfies at least one of the following conditions. We define them as the players unable to manipulate their stats.

1. Teams that played 163 or more games at the season.

Table 18: Monetary Incentives for .300 of Batting-Average: Players of the teams with close competitions

	(1)	(2)	(3)	(4)
(Intercept)	12.583 ***	-1.977	-0.539	
	(0.863)	(1.461)	(1.430)	
PA	0.004 ***	0.004 ***	0.005 ***	0.003 ***
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	-1.611	0.115	-3.148	-3.021
	(3.140)	(2.623)	(2.629)	(2.417)
AVG_300	-3.562	-1.998	-1.022	-1.451
	(2.169)	(1.947)	(1.927)	(1.686)
AVG * AVG_300	12.815	7.425	4.122	5.251
	(7.074)	(6.342)	(6.285)	(5.497)
Age		0.826 ***	0.796 ***	0.799 ***
		(0.083)	(0.080)	(0.075)
Age^2		-0.012 ***	-0.011 ***	-0.011 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	1119	1119	1114	1114
R squared	0.267	0.466	0.491	0.669
F statistic	120.826	190.300	114.724	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

The number of games is set at 162 since 1962. When some teams are lined up in winning percentage, however, then they play an additional match to decide which team to advance to the postseason games, called "One -game playoff." Needless to say, such teams went through too close competitions to adjust the players' plate-appearances to manipulate their batting-averages.

2. Teams that player 161 or less at the season.
   When a game was not played due to weather or other reasons, it is to be rescheduled. However, at the end of the season and when the standings of the team had been established, it is the case that the game is not taken place and the team end their regular season.
3. Teams that won(missed out on) the playoff berth by one game.
   Again, they experience close competition: almost at the end of the season.
4. All teams in the 1994 season.
   The regular season in 1994 was abruptly terminated on August 11th, because of the strike by the Major League Baseball Players Association. It is natural to assume that players would not predict when the season ended, so they could not manipulate their stats, either.

Figure 19 shows the result of McCrary(2008)'s manipulation test. As that of Section 5.3.3., no bunching was found ($z = .91, p = 36\%$).

Results again show no significant difference with those of previous sections. The estimated coefficients of interest are the opposite of what had been expected.

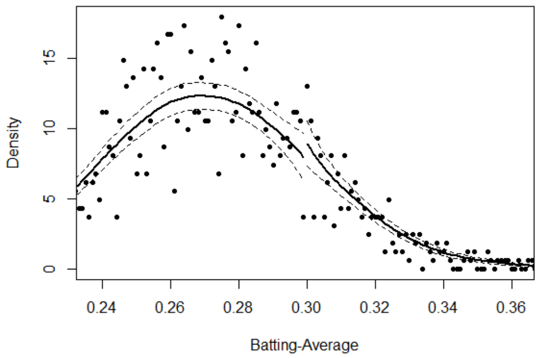

Figure 19: Manipulation Test for Players "Unable" to Manipulate Stats

In summary, we found that players did not have monetary incentives at their observed internal goals. Observing the reward functions $f(.)$ for each performance stats does not either jump or kink at the possible reference points that are reported in Section 5.1. That is, they adjust their effort level to make their performance stats just above the reference points, because of their reference-point dependent preferences. Although we do not specify which of the two characteristics that lead to bunching is more appropriate, the players' utility function can be described with such characteristics. In the next section, we consider other alternative explanations by conducting additional analysis and empirical evidence.

6 Alternative Interpretation and Some Evidence

In section 5, our analysis presented that there in fact exists bunching in some of the batting stats, but no evidence was observed in their contracts, which supports the hypothesis that these observations are driven by reference dependent preference of the players. Here we consider some possible alternatives and additional discussions about our results.

6.1 Incentivized Contract

One most possible interpretation of our results is the incentivized design of the contract. So far, we tested monetary incentives for the player, analyzing only the fixed parts of the contract. However, players often sign contracts with an additional bonus according to their performances. Even though the results in Section 5.2 did not support the existence of incentives in the fixed salary, it may occur as these additional rewards. Here we present that this hypothesis will be rejected by showing the specific contracts of some players.

Table 19 shows the specific contents of the MLB position players' contracts, quoted by *Cot's*

Baseball Contracts from *Baseball Prospectus*, an open website that discloses information about that. In addition to signing bonus, fixed payment (we analyzed this part), and other optional bonus or service, players receive some monetary incentives according to their performance. They are roughly grouped into two: award bonus and bonus for reaching a certain number of their stats. While the former includes winning Gold Glove or All-Star Game selection (a match between the two big leagues of MLB, each of which is composed by players selected by the manager and the fan's vote), the latter consists of only round numbers with plate-appearances or games they attended, neither batting-average, on-base percentage nor home runs. Besides, there are at most 2 or three position players who sign such contracts. Pitchers are more likely to agree to ones with performance bonuses, whose trigger stats are also related to attendance: the number of games appeared, or innings pitched. Therefore, we can conclude that in the additional bonus parts of their contracts, there are no incentives that encourage them to manipulate their batting-average, on-base percentage or other batting-stats.

Team managers have to design contracts with limited budget constraints. Plate-appearances given to single teams are on average constant throughout the year because they play the same number of games, so the team manager distributes the fixed numbers of plate-appearances to the players. That is, managers can predict how many players at most achieve their goals. On the other hand, it is hard for the managers to estimate the total numbers of the players that achieve some round numbers of batting-average or homerun. If almost all the players reached the benchmarks, then even if they led the team to win, managers have to owe additional expenditure. This point can be a supportive discussion of our results.

6.2 Contract Length

Skilled players often sign contracts with a plural-year duration. This is related to why we analyzed using the sample of players who held the right of free agency. Furthermore, we should also take account of their contract length, that is, until when the players are insured to play for the team they signed because it can be considered as the additional monetary bonus.

Krautmann & Oppenheimer(2002) researched this point. They used the salary dataset of MLB from the 1990 to 1994 seasons and regressed log salary on an interaction term of the performance proxy and the contract years they signed.

$$\ln(SAL_{it}) = \beta_1 + \beta_2 PERF_{it} + \beta_3(PERF_{it} * LENGTH_{it}) + \beta_4 LENGTH_{it}$$

The model is employed by Krautmann & Oppenheimer(2002). According to their results, the coefficient of the interaction term, β_3, was estimated to be negative at the 1% of significance. In words, the longer the contract years at once stretch, the less the return to their performance becomes. This is caused, they claimed, by the player's risk-aversive preference that dislikes the

Table 19: Descriptions of the Contract of the Specific Players

- Ichiro Suzuki, Outfielder, 4-year contract with Seattle Marinars (2004-'07)
  - signing bonus- $6M
  - fixed payment- 04:$5M, 05:$11M, 06:$11M, 07:$11M
  - performance bonuses- $1.25M in performance bonuses for plate appearances
    * $50,000 each for 400 PAs, 2004-06
    * $0.1M each for 500 & 600 PAs, 2004-06
    * $0.1M for 400 PAs, 2007
    * $0.2M each for 500 & 600 PAs, 2007
  - award bonuses: $50,000 each for Gold Glove, All Star selection
  - trade-protection (Veto for moving the team without his acceptance): limited no-trade clause (may block deals to 10 clubs)
  - Other
    * housing allowance: $28,000 in 2004, $29,000 in 2005, $30,000 in 2006, $31,000 in 2007
    * interpreter, trainer, transportation for spring & regular season
    * 4 annual round-trip airline tickets from Seattle to Japan
- Eric Sogard, 2nd-baseman, single-year contract with Milwaukee Brewers (2018)
  - fixed Payment- $2.4M
  - performance bonuses- : $0.15M each for 30, 50, 70, 90 games. $50,000 for 120 games
- Alex Avila, Catcher, two-year contract with Arizona Diamondbacks (2018, 2019)
  - Fixed Payment- 18:$4M, 19:$4.25M
  - annual performance bonuses: $25,000 each for 350, 400 plate appearances. $50,000 each for 450, 500 PA. $0.1M for 550 PA.

(Quoted from *Cot's Baseball Contracts*)

risk of being fired. From a viewpoint of our model, those who achieve their goals, receive an additional bonus, but instead of getting higher baseline salary, and they, therefore, choose to sign the contract with a longer duration. For the team manager, it is profitable to propose such contracts, which may enable them to hold highly skilled players with relatively reasonable costs. In these days, it is usual that players sign the plural-year package contracts with the right to opt-out: the player or the manager nullify the contract while it is under duration, for the players to get some better contract, or for the manager to modify the contract or release the player. So a more complicated model might be required to describe this situation, but it helps us to consider these nonmonetary bonuses.

6.3 By-Time Analysis

Our research used data from a wide range of time: 62 years for performance stats, 31 years for salaries. Through such a long time, techniques of the players or the quality of instruments have evolved, which leads to change in mean or the standard values of the stats: that is, unlike the reference point "par" of golf, the reference point of baseball might move over time. Besides, there may have been a lot of changes in the design of the contract they agreed. Here we consider time-variable elements in our analysis. Specifically, two main possible effects change the contract design: one is the relative market power of the players and another change in the relative importance of each performance stats.

Relative market power has a direct relation to the contract. Before the system of free agency was introduced, players were forbidden to move to other teams without permission by the team which they belong to. The 1994 strike by the Players Association of Major League Baseball, against the team owners to request improvement of their treatment, also may have a great influence on their contracts (See Appendix about the specific information about free agency and the Strike).

It is also important to capture the change in the evaluation of each stats. Through the history of baseball, a lot of stats have been invented to measure the performance/ability of the player, and it has been argued which stats are the most efficient ones to evaluate them. One of the most important revolutions was the publication of *'Moneyball'*(2003), written by Michael Lewis, a financial reporter. In this book, he described that batting-average was not appropriate to measure the quality of the players: on-base percentage is more close correlation with total runs the team earns in the season. In practice, Oakland Athletics applied this strategy to form the members of the team, and won the playoff. This story was widely spread and changed the sense of view about the baseball stats.

The impact of this publication was so great that it was evaluated in an economic article. Hakes and Sauer (2006) tested Lewis's claim in the econometric specification. They reported that on-base percentage gave us a better explanation about the winning percentage of the team than batting-average, but team managers had been taking batting-average of more importance when evaluating players. After *Moneyball* was published, however, their evaluation revolved. In 2004, a year after its publication, the estimated return to on-base percentage for the players increased, compared to the previous 4 years.

Then, one possible question occurs: "Does the tendency of manipulation/discontinuous contract design also change through the history of baseball?"

In this section, we replicate the methodologies conducted in the previous sections, but sorting the sample into periods below:

1. Before Free Agency (1957 - 1975)
2. After Free Agency and Before Strike (1976 - 1994)
3. After Strike and Before *Moneyball* (1995 - 2003)
4. After *Moneyball* (2004 - 2017)

Sample B does not include data from 1957 to 1986, so in the section of monetary incentive, We conducted tests for only three parts except for "Before Free Agency." From here, we focus on the three important batting stats: batting-average, on-base percentage, and homerun.

6.3.1 Bunching

Table 20 shows the results of the McCrary (2008)'s manipulation tests, for each grouped samples. Compared to the full-sample analysis conducted in Section 5.1, we observed partly different results for each stats.

First of all, .300 of batting-average is the most solid benchmarks of the players. Each subsample shows the bunching at the cutoff point, at least at the 5% level of significance. Although there exist skeptical discussions on the validity of batting-average as a performance stats, it still works as an important status for the position players. There are no other stats that show such a consistent tendency among the samples.

On the other hand, .250 seems not to be regarded as important after *Moneyball*, as other previous days. In this term, the bunching at .250 becomes insignificant. Compared to the samples of the old days, the average level of the batting-average has been increasing. The mean value of each samples are .259 (samples of '57-'75), .264 ('76-'94), .271 ('95-2003), .263 (2004-2018), respectively, median values of which are almost the same. Note that in fact, restricting the sample to the years until 1965, then the manipulation test shows no statistical significance at .300 ($z = 1.577, p = 11.45\%$). So as we mentioned in Section 5.1, it may be because we extended our data range to 1957, that .250 works as a reference point. For samples before 1965, bunching at .250 was significant at 5% level ($z = 2.12$).

Excess mass of homerun was significant only in the latest subsample. After the publication of *Moneyball*, SABRmetrics argue that extra-base-hits (double, triple and home run) have a significant effect on an increase in the runs their team earns (Information of SABRmetrics is provided in Appendix). Besides, on-base percentage, surprisingly, showed no evidence for bunching in the subsample level. Thus, as we mentioned above, players consider these stats as less important ones to evaluate players than .300 of batting-average. Otherwise, we have to consider the possibility that there may have existed some different design of contracts. In the next section, we describe an analysis of these results, showing that there also did not exist any monetary incentive to achieve these points.

Table 20: Bunching Test for the Grouped Sample by Time

stats, cutpoint		'57-'75	'76-'94	'95-2003	2004-	full sample
AVG, .300	bw	.023	.020	.022	.019	.019
	θ	.573	.566	.310	.403	.499
		(.146)	(.120)	(.130)	(.120)	(.067)
	z	3.934***	4.732***	2.393*	3.376***	7.442***
AVG, .250	bw	.028	.028	.032	.027	.024
	θ	.250	.151	.306	.121	.212
		(.080)	(.069)	(.094)	(.076)	(.042)
	z	3.149**	2.188*	3.242**	1.595	5.061***
OBP, .350	bw	.031	.030	.036	.030	.024
	θ	.137	.149	-.035	.137	.139
		(.089)	(.081)	(.093)	(.082)	(.049)
	z	1.538	1.846	-.380	1.672	2.854**
HR, 20	bw	6.313	6.677	10.165	7.273	5.309
	θ	.222	.214	.145	.315	.259
		(.150)	(.123)	(.129)	(.112)	(.075)
	z	1.479	1.751	1.117	2.819**	3.465***
Note		***: $p < 0.1\%$, **: $p < 1\%$, *: $p < 5\%$. Bandwidth is optimized following the method of McCrary(2008).				

6.3.2 Monetary Incentive

Table 21 shows the results of the local-linear regression conducted in Section 5.2 for the restricted samples. Each column shows the statistics for the subsamples, and the far-right edge column displays the result using the full sample, the same one as Table 4. Each result includes the other controls (that is, results correspond to "Yes" in "other control" of Table 4). Results of the analysis including the interaction terms are shown in Table 24 to 25 (for .300 of batting-average. Analysis for other cutoffs show the same results). According to these results, they are consistent with the analysis in Section 5.2 for all the stats: Any notch or kink in the players' monetary rewards was not observed. Additionally, we describe the same analysis for the players with free agency in Table 22 that again shows the same results. All the statistics that stand for the notch are insignificant to their fixed part of the salaries. Therefore, we repeat the same conclusion as Section 5: there are no monetary incentives that verify the reason that players bunch their stats. It implies that the players have preferences that lead to excess mass, or that the utility functions of the players have kink or notch at each cutoff point, which works as a reference point for the player.

In sum, we obtain the additional conclusion of this section: there may have some factors that cause a change in the players' attitude to their stats about bunching. On the other hand, it is consistent over time that team managers do not make the reward function $f(.)$ that leads players to bunching. That is, it is caused by the players' reference point dependent preferences,

Table 21: Local-Linear Regression for the Grouped Sample by Time

stats, cutpoint	bw, type		'87-'94	'95-2003	2004-	full sample
AVG, .300	Optimal	bw	.024	.042	.030	.045
		Obs.	697	1806	1872	5930
		estimate	-.034	.064	.066	.034
			(.137)	(.092)	(.103)	(.056)
		z	-.250	.697	.637	.615
AVG, .250	Optimal	bw	.036	.043	.075	.048
		Obs.	1482	1806	3991	7271
		estimate	.154	.064	.076	.070
			(.084)	(.092)	(.060)	(.052)
		z	1.825	.697	1.277	1.340
HR, 20	Optimal	bw	4.183	3.685	2.46	3.30
		Obs.	341	371	475	1307
		estimate	-.255	-.348	.343	-.002
			(.228)	(.218)	(.264)	(.141)
		z	-1.122	-1.600	1.300	-.015
OBP, .350	Optimal	bw	.031	.025	.027	.045
		Obs.	1098	1281	2042	6525
		estimate	.109	-.151	-.030	-.013
			(.106)	(.120)	(.093)	(.049)
		z	1.031	-1.262	-.323	-.272

Note: ***: $p < 0.1\%$, **: $p < 1\%$, *: $p < 5\%$.
Bandwidth is optimized following the method of Imbens-Kalyanaraman (2009).

supporting Pope and Simonsohn (2011)'s interpretation.

7 Conclusion

This paper has considered the possible existence of monetary incentives behind the behavior that appears to be related to reference-point dependent preference.

Three important findings are obtained. First, MLB position players adjust their effort level so that their performance stats to achieve their internal goals. These internal goals are round numbers, but not all of them work as such benchmarks. As a result, around the cutoff points, excess mass or bunching are observed. Second, at least in the fixed part of their contracts, there is no clear evidence that supports the existence of monetary incentives. Although there remains a room for discussion in stolen-bases, the salary scheme did not kink or jump at the cutoff points where bunching was observed. That is, players try to reach the benchmark of each performance stats, even though there are not any monetary incentives that lead them to do so. Thus, these observed efforts are likely to be caused by their reference-point dependent preferences, as Pope and Simonsohn argued. Finally, repeating the analysis with subsamples divided by time, the players' attitude to the cutoff points has been changing over time, and it results in the change in

Table 22: Local-Linear Regression for the Grouped Sample Including FA Players

stats, cutpoint	bw, type		'87-'94	'95-2003	2004-	full sample
AVG, .300	Optimal	bw	.060	.032	.039	.026
		Obs.	218	229	354	509
		estimate	-.026	-.309	-.186	-.253
			(.247)	(.182)	(.182)	(.138)
		z	-.108	-1.700	-1.020	-1.832
AVG, .250	Optimal	bw	.018	.023	.078	.058
		Obs.	123	227	716	1367
		estimate	.425	.293	.047	.084
			(.281)	(.230)	(.103)	(.082)
		z	1.512	1.272	-.448	1.020
HR, 20	Optimal	bw	5.35	3.504	3.566	3.50
		Obs.	47	70	102	206
		estimate	.004	-.701	-.337	-.273
			(.284)	(.492)	(.513)	(.296)
		z	-1.600	-1.423	-.657	-.924
OBP, .350	Optimal	bw	.034	.042	.031	.043
		Obs.	154	344	395	1044
		estimate	.080	-.174	.115	.021
			(.291)	(.179)	(.188)	(.107)
		z	.276	-.971	.616	.196

Note: ***: $p < 0.1\%$, **: $p < 1\%$, *: $p < 5\%$.
Bandwidth is optimized following the method of Imbens-Kalyanaraman (2009).

Table 23: Before Strike

	(1)	(2)	(3)	(4)
(Intercept)	12.995 ***	-1.050	0.524	
	(1.807)	(2.009)	(2.032)	
PA	0.003 ***	0.003 ***	0.002 **	0.003 ***
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	-3.036	1.653	-2.244	-2.713
	(6.324)	(5.534)	(5.580)	(5.109)
AVG_300	-4.805	-3.749	-4.037	-0.368
	(3.422)	(2.961)	(2.908)	(2.499)
AVG * AVG_300	16.754	12.832	13.842	1.628
	(11.329)	(9.808)	(9.641)	(8.317)
Age		0.769 ***	0.744 ***	0.683 ***
		(0.088)	(0.086)	(0.080)
Age^2		-0.011 ***	-0.011 ***	-0.010 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	703	703	697	697
R squared	0.214	0.409	0.432	0.631
F statistic	64.346	99.546	57.617	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 24: After Strike, Before 'Moneyball'

	(1)	(2)	(3)	(4)
(Intercept)	11.421 ***	-4.372 ***	-2.434 *	
	(0.663)	(1.091)	(1.092)	
PA	0.004 ***	0.004 ***	0.003 ***	0.002 ***
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	2.574	2.412	-2.747	-1.072
	(2.428)	(2.078)	(2.070)	(2.021)
AVG_300	-1.630	-0.863	-0.684	-0.492
	(1.482)	(1.323)	(1.304)	(1.271)
AVG * AVG_300	5.779	3.288	2.691	1.895
	(4.859)	(4.332)	(4.270)	(4.157)
Age		0.968 ***	0.944 ***	0.972 ***
		(0.060)	(0.060)	(0.059)
Age^2		-0.014 ***	-0.014 ***	-0.014 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	1782	1782	1771	1771
R squared	0.320	0.515	0.554	0.606
F statistic	226.726	360.087	246.146	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

Table 25: After 'Moneyball'

	(1)	(2)	(3)	(4)
(Intercept)	13.262 ***	-2.952 *	-1.635	
	(1.056)	(1.229)	(1.213)	
PA	0.004 ***	0.004 ***	0.003 ***	0.002 ***
	(0.000)	(0.000)	(0.001)	(0.001)
AVG	-1.601	1.583	-2.049	-0.004
	(3.742)	(3.031)	(2.981)	(2.924)
AVG_300	-0.990	-1.307	0.113	0.212
	(2.142)	(1.752)	(1.703)	(1.644)
AVG * AVG_300	3.990	4.638	-0.132	-0.539
	(7.079)	(5.784)	(5.620)	(5.425)
Age		0.922 ***	0.912 ***	0.942 ***
		(0.059)	(0.057)	(0.056)
Age^2		-0.013 ***	-0.013 ***	-0.013 ***
		(0.001)	(0.001)	(0.001)
Other Controls			X	X
Fixed effects				Season, Team, Position
#observations	1880	1880	1872	1872
R squared	0.217	0.476	0.511	0.585
F statistic	143.016	338.800	224.065	
P value	0.000	0.000	0.000	

Linear regressions on loggarithm of players' salary next season.
Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Robust standard errors are in the parentheses. The bandwidth in each sample is chosen following to Imbens and Kalyamanaran (2009).

the tendency of bunching. Among the argued cutoffs, however, .300 of batting-average remained consistent bunching over time. Thus, we conclude that it works as a solid benchmark for the players. On the other hand, there are no monetary incentives in the reward functions for all the subsamples. This result supports the second finding.

Our results indicate that professional sports players seem to have preferences that yield utility not only by the monetary rewards but also by their performance situationally. It may suggest that dealing with the cases of professional occupation requires some assumption that workers do not consider the time for leisure and works as a tradeoff. For the team managers, on the other hand, our analyses help make better contract packages that attract the skilled players more, or search the player relatively underestimated, to get the players more "efficiently." We should pay attention to the opposite-side approach: where kink or notch observed about the reward function, even though any bunching does not exist. Although players do not feel important as their reference points, there may be some cutoff points that the team managers regard as reference points for evaluating players.

Simultaneously, as we described, we are required to continue the analysis of monetary incentives for performance stats. In Section 6.1, we suggested that contract durations be substituted into direct monetary rewards. Rewards may work in the stage before signing contracts: players who achieve the cutoff points may be likely to be offered by MLB teams than those who do not. Furthermore, although we do not include them in this paper, we might obtain another implication by considering the rewards they may receive after their retirement: "reaching .300 of batting-average 4 times in his career," should be some outstanding signals to get jobs as a coach of the baseball team, or a commentator in the TV show.

We can apply our results to design a more efficient contract in the situation of labor economics. The relative importance of the stats with expected runs or winning-percentages differs from one another. For example, On-base percentage is more closely correlated to them than batting-average. Besides, the number of stolen-base has little correlation with them (the correlation is as weak as $r = 0.07$). For the team managers, then, players' too much adherence to the stats that do not affect winning-percentage or ones that may be tradeoff-relation with more important stats (for example, batting-average and on-base percentage). To prevent such inefficiency, they are sometimes required to make contracts that lead players to pay more attention to the prior stats. Of course, our analysis presents some implications for them. This viewpoint can also be availed in other contract design.

As we mentioned above, in MLB, the performance stats have been recorded for almost 150 years, and thanks to the community of the fans, we can fairly compare the players played in the different generations. As we used in this paper, rich information about their contracts has been published. There exist only a few studies for international comparison: Many countries around the world, Japan, Germany, Italy or Australia and so on, have their professional leagues. Of course, there are several amateur players, those who do not receive any reward for their plays, there also exists room for comparison among them.

To conclude our paper, we state that it is worth continuing analysis of baseball, both for the sports itself and economics.

Appendix

A. Theoretical Frameworks

Many sports are so rich in performance stats that record plays in variable viewpoints. Among them, especially, baseball performance stats can distribute individual-separable evaluation to the players, as well as units of the whole team. Thus, baseball performance stats are interpreted as reliable proxies for the skills of the players. Here we assume that the players are trying to maximize them, subject to their internal costs depending on their talented skills. After observing the stats the players, team managers evaluate them and propose contracts of the next year (they also include the player-specific characteristics: age, position and so on). Then, the players' monetary rewards are given by the following benefit function:

$$f_{it+1} = f(X_{it}, Z_{it}),$$

where F stands for the monetary reward function to the player i at time $t+1$, and X_{it} and Z_{it} express the value of the stats and other player-specific characteristics, respectively.

We assume that X_{it} is determined by the effort level of each player. To increase the effort level, they have to pay additional intrinsic costs. Let e_{it} be an effort level of the player i at season t, and $c_{it}(.)$ be an continuous, increasing and convex intrinsic cost function. Note that the required cost to achieve the same level of X varies by players, according to their skills. Therefore, the maximization problem of the players are described as follows:

$$\max_{e_{it}} U_{it} = f(X(e_{it}), Z_{it}) - c_{it}(e_{it}).$$

We assume the additive separable function of the utility.

An alternative specification is that each player evaluates himself as an athlete: he directly yields benefit by his value of the performance stats. That is, players decide according to their benefit function

$$b_{it} = b(X_{it}, Z_{it}).$$

In this model, the player no longer pays attention to his future monetary rewards. In this model, we again assume the intrinsic cost determined by the player's skill and effort level. Another specification of the maximization problem is described as follows:

$$\max_{e_{it}} U_{it} = b(X(e_{it}), Z_{it}) - c_{it}(e_{it}).$$

These two ways of the specification are different only in the first terms of the utility function $U(.)$.

Then, we consider the functional features of $f(.)$ or $u(.)$, which cause bunching of the histograms of the stats. In this paper, we suppose $f(.)$ or $b(.)$ satisfy at least one of the two assumptions quoted in Section 2: one is "kink" at the reference point, and the other is "notch" of the function.

1. **"notch" at** r (Figure 20)
   The first possible assumption is a discontinuity at a certain cutoff point of the function, described as follows:

$$\lim_{\epsilon\to 0} b_r(r+\epsilon) \neq \lim_{\epsilon\to 0} b_r(r-\epsilon)$$
$$\lim_{\epsilon\to 0} f_r(r+\epsilon) \neq \lim_{\epsilon\to 0} f_r(r-\epsilon)$$

   This is the discontinuous form of the function at the cutoff point r, introduced by Diecidue and Van de Ven (2008).
2. **"kink" at** r (Figure 21)

$$\lim_{\epsilon\to 0} b'_r(r+\epsilon) \neq \lim_{\epsilon\to 0} b'_r(r-\epsilon)$$
$$\lim_{\epsilon\to 0} f'_r(r+\epsilon) \neq \lim_{\epsilon\to 0} f'_r(r-\epsilon)$$

   $u'(.)$ and $f'(.)$ stand for the first-order differential of $u(.)$ and $f(.)$, respectively ($u(.)$ and $f(.)$ are assumed to be continuous differential). As we explained in Section 2, this is the primitive form of reference dependence, introduced in Kahneman and Tversky (1979) and Tversky and Kahneman (1992).

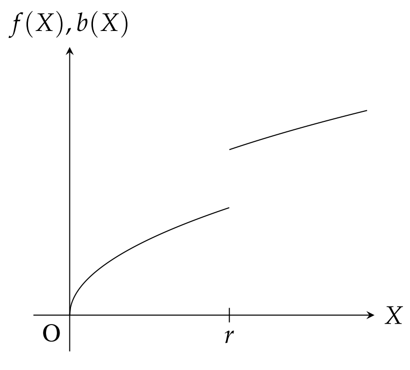

Figure 20: "notch" at the reference point

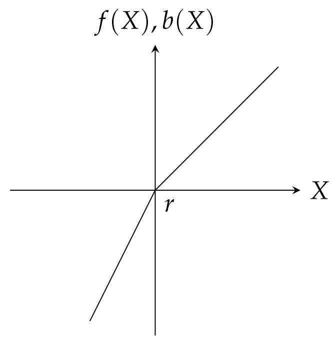

Figure 21: "kink" at the reference point

Allen *et al.* (2016) also utilized these types of model, and specified bunching of the marathon runners' finish time around the round numbers.

We cannot determine which model is appropriate to interpret the observed behavior of the players: bunching around the possible reference points. Our interest is that which of the two assumptions are observed. If the players' reference-dependent preferences for the performance stats cause bunching, then $b(.)$ has functional features that represent reference dependence. On the other hand, if their contracts are designed to lead them to the bunching around some possible reference point, now we can regard that $f(.)$ has the features described.

Both two possible assumptions result in bunching at the cutoff point: in our paper, we consider that these cutoffs are the round numbers of the performance stats, such as .300 of batting-average. [3] In words, the players' salary scheme might be designed such that players make an effort to meet their cutoff points. In the rest of this subsection, we explain the possible design of the contracts.

Let us consider the first case: the monetary reward "jumps" at the possible reference point. Then, the salary function is decomposed into two terms as follows.

$$f_{it+1}(.) = a(X_{it}, Z_{it}) + D_{it}Bonus(X_{it})$$

$f(.)$ is determined by the number of the performance stats X and other player-specific characteristics Z, note that this term is continuous in X. D is the dummy variable that depends on X: to indicate if the player's stats are above the cutoff point, but otherwise zero. If the players achieve the cutoff point like .300 of batting-average, then f discontinuously rises by the appearance of *Bonus* term. And as a result, the salary scheme with this form of function causes the excess mass around the cutoff, by the players who desire the additional monetary rewards.

The second case, on the other hand, expresses the "kink" of f as follows:

$$f_{it+1}(.) = (1 - I_{it})g(X_{it}, Z_{it}) + I_{it}h(X_{it}, Z_{it})$$

I is the same indicator as D, which indicates 1 if X is larger than the cutoff, but 0 otherwise. Two salary schemes are depending on the players' performance stats, represented by g and h. g is applied for the players with the stats below the cutoff, and h is for those with above there. Furthermore, h consists of two terms: $h(.,.) = g(c, Zit) + k(X_{it} - c)$. c stands for the given cutoff. In words, players receive the continuous reward in their performance stats, but the return to the stats changes after they achieve the cutoff: in our assumption, decrease. More directly, the marginal rate of return to their performance stats discontinuously diminishes: kinks at the reference point.

While we cannot observe b_{it} directly, f_{it} can be observed as the contracts that players agreed. If f_{it} has the discussed functional features, then we conclude that the contracts' contribution to bunching. In the next section, we describe how to specify them.

[3] Therefore, it is not so important to determine which of the two is appropriate to describe the bunching itself: that is, our interest is whether $f(.)$ has these types of functional features in the number of performance stats.

B. Reference to the stats

Here we describe the details of the stats used in the paper.

- Definition
  - Batting-Average (AVG)

$$\text{AVG} = \frac{\text{Base-Hit}}{\text{At-Bat}}$$

    The rate of base-hit. At-Bat(AB) is calculated as AB $=$ PA $-$ Base-on-Balls $-$ Hit-by Pitch $-$ Sacrifice-Bunt $-$ Catcher-Interferance.
    AVG depends only on the number of base-hit, so when players intend to manipulate AVG, they try to get base-hit, not base-on-balls. Moreover, AVG does not identify the number of bases they get at one base-hit: single, double, triple, and a home run.
  - On-Base Percentage (OBP)

$$\text{OBP} = \frac{\text{Base-Hit} + \text{Base-on-Balls} + \text{Hit-by-Pitch}}{\text{PA} - \text{Sacrifice-Bunt} - \text{Catcher-Interferance}}$$

    OBP is different from AVG in that it takes base-on-balls and hit-by-pitch into account. *Moneyball* and Society of American Baseball Research (SABR) argues that OBP is more correlated to the runs the team earns than AVG.
  - Batting, Fielding, BaseRun
    SABRmetrics, which considers baseball by the scientific approaches, have argued that which player is more important for the team. You measure Players' contribution in the following procedure:
    1. Convert each play (base-hit, strike-out, sacrifice-bunt, . . .) into runs-value.
    2. Convert runs-value into a contribution to the number of wins the team obtains from the contribution of the player.

    Batting, Fielding, and BaseRun stand for the first step of the identification: the estimated number of runs the player produces by his batting, fielding, and base-running above the average. For the team they play for, they are kinds of clear indicators of the players' contribution.
  - WPA (Win-Probability-Added)
    By the statistics of the games, SABRmetrics have obtained the winning percentage at each specific situation of the game: scores (2-runs forward, 3-runs behind and so on) inning, runners on-base. Then, we can define the change in the win-percentage generated by the single play. WPA is a cumulative stats that stands for the summation of this change that a single-player produced.
    Most studies by SABRmetrics evaluate only the result of each play, excluding the context then: for example, if players hit a home run, then we treat it equally. That is, it does not matter the home run was a walk-off homer occurred when the game was tied, or solo-homer when the game had been broken. This is because this type of skill called "clutch" does not correlate between a year to year.
    However, it may be the case that team managers evaluate this type of skills in the contracts, so our regression analysis for monetary incentives (Section 5.2 and 6.3.2), we employed WPA as a regressor. We include the stats, distinguishing positive (raises the winning-percentages)

plays and negative (decline) ones, and each value divided by the plate-appearance. And actually, the results of each regression supported that these terms affect the monetary rewards.

C. Important Events Related to Section 6.4

- Free Agency
  Free agency, the right for players to negotiate and choose any teams he wants, was regulated in 1975. Until then, players were forced to play for the team serving, except for the manager fired him. Thanks to this rule, now any players who have been serving in MLB for about 6 seasons, are eligible to get a free agent. By this procedure, they are free from the monopsony contract with the team playing for. And so, players got more likely to be evaluated accurately with his actual skills. The right to arbitration is the same kind of this.
- Strike by the Players Association (1994)
  In 1994, the Players Association of MLB declared to make Strike, to stop the owners' cartel. These days, because of the increased salary of the skilled players, they are suffered from keeping their payroll in their budget constraint. So to resist this, they are tried to make cartels that refrain from proposing any or sufficient contract to the super-skilled players. This results in such cases that players who have enough skills chose to go to Nippon Professional Baseball (NPB).

References

[1] Abeler, Falk, Goette, and Huffman. 2011 Reference Points and Effort Provision *American Economic Review,* VOL. 101, NO. 2, APRIL 2011 (pp. 470-92)

[2] Allen, Dechow, Pope and Wu. 2016. Reference-Dependent Preferences: Evidence from Marathon Runners *Management Science* 63(6):1657-1672.

[3] Bandiera, Barankay, and Rasul. 2007. Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment *The Quarterly Journal of Economics,* Volume 122, Issue 2, May 2007, Pages 729–773

[4] Berger and Pope. 2011. Can Losing Lead to Winning? *Management Science,* Volume 57, Issue 5

[5] Crawford and Meng. 2011. New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income *American Economic Review*, VOL. 101, NO. 5, AUGUST 2011 (pp. 1912-32)

[6] Ericson and Fuster. 2011. Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments *The Quarterly Journal of Economics,* Volume 126, Issue 4, November 2011, Pages 1879–1907

[7] Gill and Prowse. 2012 A Structural Analysis of Disappointment Aversion in a Real Effort Competition *American Economic Review*, VOL. 102, NO. 1, FEBRUARY 2012 (pp. 469-503)

[8] Hakes and Sauer. 2006. An Economic Evaluation of the Moneyball Hypothesis *Journal of Economic Perspectives* Volume 20, Number 3—Summer 2006—Pages 173185

[9] Imbens and Kalyanaraman. 2009. *NBER Working Paper Series.* 14726

[10] Kahneman and Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* Journal of the Econometric Society47 (2):263291.

[11] Krautmann and Oppenheimer. 2002. Contract Length and the Return to Performance in Major League Baseball *Journal of Sports Economics* February 2002

[12] Lee and Lemieux. 2010. Regression Discontinuity Designs in Economics *Journal of Economic Literature* 48 (June 2010): 281–355

[13] McCrary. 2007. Manipulation of the running variable in the regression discontinuity design: A density test *Journal of Econometrics* 142 (2008) 698–714

[14] Paarsch and Shearer. 2001. Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records *International Economic Review,* Volume41, Issue1 February 2000 Pages 59-92

[15] Pope and Schweizer. 2011. Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes *American Economic Review* 101 (February 2011): 129157

[16] Pope and Simonsohn. 2011. Round Numbers as Goals: Evidence From Baseball, SAT Takers, and the Lab *Psychological Science* 22(1) 7179

[17] Rees-Jones. 2018. Quantifying Loss-Averse Tax Manipulation *Review of Economic Studies* (2018) 85, 1251–1278

[18] Scully. 1974. Pay and Performance in Major League Baseball *The American Economic Review* Vol. 64, No. 6 (Dec., 1974), pp. 915-930

[19] Shearer. 2004. Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment *The Review of Economic Studies*, Volume 71, Issue 2, April 2004, Pages 513–534

[20] Tversky and Kahneman. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty *Journal of Risk and Uncertainty*, 5:297-323 (1992)