Southern Economic Journal 2008, 75(2), 574–590

Exit Discrimination in Major League Baseball: 1990–2004

Peter A. Groothuis* and James Richard Hill†

Using a panel study of annual Major League Baseball (MLB) data (1990–2004), we do not find evidence that race affects the career duration of black hitters. Our findings are inconsistent with results from a study by Jiobu (1988) who used 1971–1985 data that found that race decreased career length, *ceteris paribus*, for black hitters but not Hispanics. The difference in results could be due to our use of seasonal-variant data; Jiobu used only career statistics in his research. Another interpretation of the differing results is that market competition overcame discrimination. This could be interpreted as an affirmation of Becker's theoretical work on discrimination. Our results for pitchers also do not indicate any exit discrimination against the minority groups.

JEL Classification: J71, L83

1. Introduction

The topic of discrimination in labor markets has attracted a great deal of research by economists over time. From the theoretical constructs developed by Becker (1971) to the development of the residual method by Oaxaca (1973), economists have struggled to determine the cause and effects of prejudice on minority workers. Professional sports have provided a fertile field for researchers of this topic because of the abundance of performance data and salary information available. In Kahn's (1991) article on discrimination in professional sports, he reviews empirical findings from numerous articles published between 1972 and 1989. He divides his summaries under four major headings: salary discrimination, positional discrimination, customer discrimination, and hiring discrimination. Many of the articles offered findings on more than one of the topic headings.

Salary discrimination drew the most attention. Twenty-four articles with empirical work on this topic are critiqued by Kahn (1991): 12 on baseball, eight on basketball, and two each on football and hockey. The strongest evidence of salary discrimination against black athletes comes from studies of the National Basketball Association (NBA) based on data from 1984–1986. However, more recent research (Bodvarsson and Brastow 1999; Hill 2004) has failed to find salary discrimination in the NBA using data from the 1990s. Twelve articles involving

* Department of Economics, 3093 Raley Hall, Appalachian State University, Boone, NC 28608, USA; E-mail groothuispa@appstate.edu.

† Department of Economics, 317 Sloan Hall, Central Michigan University, Mt. Pleasant, MI 48859, USA; E-mail hill1jr@cmich.edu; corresponding author.

We would like to thank the audience and discussants at the 2006 Southern Economics Association annual meeting for helpful comments. This paper has been greatly improved by the comments and suggestions of three anonymous referees. All remaining errors are our own.

Received January 2007; accepted September 2007.

research on positional segregation were reviewed (Kahn 1991): five from baseball, three from football, two from basketball, and two from hockey. Evidence suggested positional segregation or stacking existed for blacks in baseball and football; evidence also suggested that blacks were underrepresented at the center and forward positions in the NBA. Customer discrimination was the main or auxiliary focus in 15 of the papers (Kahn 1991): nine from basketball and six from baseball. Results are mixed for attendance at NBA games; a recent article (Kanazawa and Funk 2001) finds a negative correlation between the percentage of blacks on the roster and cable TV ratings. In baseball, evidence from mid-1970s revenue data fails to find customer discrimination, but studies using earlier data did. Customer discrimination also appears to exist in the baseball trading card market (Nardinelli and Simon 1990).

Of the 13 pieces of empirical work focused on hiring discrimination reviewed by Kahn (1991), six deal with data from baseball, four with basketball, two with hockey, and one with football. Findings failed to confirm hiring barriers for black athletes in basketball and baseball; whites with inferior performances were not hired over superior black athletes. However, Johnson and Marple (1973) found evidence from 1970–1971 NBA data that marginal white players had longer careers than marginal black players. Hoang and Rascher (1999) explored the concept of racially based retention barriers in the NBA in more detail using data from the 1980s. They, too, found evidence that, holding performance constant, there was "exit discrimination" in the NBA. Groothuis and Hill (2004) failed to confirm Hoang and Rascher's results using more recent data, height as an added explanatory variable, and a duration model that allowed for both stock and flow samples.

Jiobu (1988) found evidence that race decreased career length, *ceteris paribus*, for black players but not Hispanics using Major League Baseball (MLB) data from 1971–1985. The purpose of this article is to determine if Jiobu's findings can be replicated using more recent baseball data and a richer duration model that has season-variant data; Jiobu used only unchanging, career performance variables in his model. Exit discrimination is more subtle than positional segregation or pay discrimination. If exit discrimination in MLB is no longer present, it would seem to affirm Becker's theoretical implications that market competition will eliminate discrimination over time.

2. Theory

Discrimination in the labor market implies that certain individuals or groups of workers are somehow treated differently than others unrelated to ability or performance; in the literature on professional sports, the focus has been on differential treatment of black, Hispanic, and French-Canadian players. The differential treatment can occur in a variety of formats. When pay discrimination occurs, workers in the group receiving the unequal treatment receive lower pay than others for the same performance level. Most researchers believe pay discrimination against black players in professional baseball was eventually eliminated by free agency and salary arbitration in the mid-1970s (Hill and Spellman 1984; Cymrot 1985). Research by Johnson (1992) even suggests that white players earn less than black players of similar ability on teams that have fewer black players.

Hoang and Rascher (1999) define exit discrimination as "the involuntary dismissal of workers based on the preferences of employers, coworkers, or customers" (pp. 69–70). Jiobu

(1988) and Hoang and Rascher (1999) concluded that career lengths for black players in MLB and the NBA, respectively, were lower than those of their white counterparts, *ceteris paribus*. While Jiobu does not make any calculations on the impact of exit discrimination on career earnings, Hoang and Rascher conclude that this form of discrimination led to an almost two and a half times greater decrease in black career pay compared with the more heavily analyzed form of pay discrimination.

Becker (1971) suggests that the source of the personal prejudice that leads to various forms of labor market discrimination may be employers, coworkers, or customers. Recent research by Kahn and Sherer (1988) on pay discrimination in the NBA and Hoang and Rascher (1999) on exit discrimination in the NBA has focused on customers as the source of the prejudice; both studies found empirical evidence to support the existence of customer discrimination. The pay premium for white players was explained by the higher value of their performances compared with black players because of the prejudiced preferences of white, majority fans.

Hoang and Rascher (1999, p. 74) hypothesized: "To satisfy the fans, there is a minimum number of white players on a team. The second assumption, that the pool of quality available talent is becoming increasingly black, causes annual replacement of players with rookies to occur mostly among black players. The white players have longer careers simply because there are fewer qualified white rookies to replace them..."

In his study of exit discrimination in MLB, Jiobu (1988, p. 532) does not specifically test for customer discrimination, but he does state: "Perhaps, motivated by the concern that white fans will not support a predominantly black team, management has silently placed an 'invisible ceiling' on the black percentage. When coupled with the desire to have a winning team, this ceiling would generate strong pressures to (a) employ as many black players as possible in order to capitalize on their performance, but (b) in order to remain under the ceiling, to eliminate black players as soon as their performance declined, and (c) to retain white players of declining but similar ability."

Research on this topic assumes that all turnover is involuntary; Kahn (1991) argues that the high salaries paid in sports make voluntary quits unlikely. Thus, these studies are essentially survival models. If white players have longer careers than black players with similar performance statistics, then exit discrimination is said to exist.

3. Data

Our data include all individuals, both hitters and pitchers, who participated in MLB from 1990 through 2004 for a 15-year panel consisting of 3185 players.[1] Jiobu's (1988) study on exit discrimination in baseball only looked at hitters. To capture the overall length of players' baseball careers, our data consists of both stock and flow samples. A stock sample consists of all ongoing careers at the start of the panel in 1990. These left-censored data are easily included because we know how many years each player had played in the major league prior to 1990.

[1] Because MLB roster limits are expanded from 25 to 40 from September 1 until the end of the season, we chose to ignore hitter observations in which the player appeared in less than 20 games in a season and pitcher observations in which the player appeared in less than six games in a season. We wanted to limit the number of players from the sample who never truly had a career in MLB but simply made the expanded roster for two or three seasons. Jiobu (1988) only included hitters in his sample who appeared in 50 or more games during their careers.

Table 1A. Hitter Panel Data (1990–2004): Means of Variables

Variables	White	Black	Foreign-Born Hispanics
Age	29.55	29.82	27.97
Slugging average	0.394	0.404	0.395
Stolen base percentage	3.70	10.61	6.17
Runs batted in per game	0.369	0.386	0.385
Home runs per game	0.746	0.805	0.804
Walks per game	0.295	0.313	0.261
Strikeouts per game	0.540	0.552	0.549
Hit by pitch per game	0.028	0.024	0.027
Sacrifice flies per game	0.026	0.025	0.025
Runs scored per game	0.378	0.441	0.407
Games	93.02	102.54	98.52
Infield	0.397	0.204	0.493
Catcher	0.221	0.016	0.142
First base	0.125	0.088	0.065
Primary DH	0.022	0.036	0.027
Number observations	3795	1621	1462
Number of players	821	283	312

Our stock sample consists of 690 players who had an average tenure of six years as they entered the 1990 season. Including a stock sample captures information on players whose careers are longer than the panel data set. Using only stock data, however, would underrepresent short-career players, so we also include flow data.

A flow sample includes all careers that start between 1990 and 2004. This sample captures many short careers in MLB. Including only flow data, however, would allow for no careers longer than 15 years, which is the length of our panel. As with most panels, our data are also right-censored where many careers were ongoing when our sample ended in 2004. Our right-censored data include both stock and flow observations. To estimate a duration model of stock and flow data, we use a technique developed by Berger and Black (1998).

The variables in our data include both season-variant and season-invariant data. Jiobu (1988) used only season-invariant data for his analysis. We report the means in Table 1A for hitters and Table 1B for pitchers. Our season-invariant data include dummy variables for player's race; a dummy variable for infielders, first basemen, and catchers; and a dummy variable for pitcher's throwing hand.[2] Our season-variant data for hitters are age, age squared, a dummy variable for each team's primary designated hitter, and performance data that include games played as well as slugging average, home runs per game, stolen base percentage, runs batted in per game, walks per game, strikeouts per game, hit by pitch per game, sacrifice flies per game, and runs scored per game. Our season-variant data for pitchers are age, age squared, and performance data that include games played as well as earned run average, wins, losses, saves, strikeout-to-walk ratio, and outs in innings pitched. In Tables 1A and 1B we report variable means by race for hitters and pitchers, respectively. To determine if exit discrimination exists in the major leagues, we analyze the data using both nonparametric and semiparametric techniques.

[2] There are race dummy variables for U.S.-born black players, foreign-born Hispanic players, and foreign-born Asian players. United States-born Hispanic players are included in the white category. This is the same approach used by Jiobu.

Table 1B. Pitcher Panel Data (1990–2004): Means of Variables

Variables	White	Black	Foreign-Born Hispanics
Age	29.08	28.44	27.61
Earned run average	4.731	4.848	4.775
Wins	4.793	4.789	4.891
Losses	4.758	4.562	4.763
Saves	2.316	3.117	2.834
Strikeout-to-walk ratio	1.903	1.723	1.940
Outs in innings pitched	256.226	248.789	256.969
Games	32.49	34.17	34.08
Left-handed	0.311	0.227	0.197
Number of observations	5389	294	1008
Number of players	1333	66	238

4. Model and Empirical Results

Nonparametric Estimates of Career Duration

To help understand career duration in MLB, we calculate yearly hazard functions as

$$h_t = d_t/n_t, \tag{1}$$

where d_t is the number of players who end their career in year t, and n_t is the number of players at risk of ending their career in year t. The hazard rate can be interpreted as the percentage of players who exited MLB given that they have survived up to some level of tenure.

Empirical Results

In Table 2A, we report the hazard rate for foreign-born Hispanic, black, and white hitters, and in Table 2B, we report the hazard rate for pitchers by race. In Figures 1A and 1B, we plot the hazard functions for hitters and pitchers, respectively, by race. All three plots show that the hazard rate gradually declines for the first seven years of tenure in MLB and then gradually climbs for the remainder of the years. In addition, both hazard plots follow an interesting pattern that resembles a U shape. In the engineering literature on useful life, this hazard function shape has become known as the "bathtub" plot (Bolla 2002). We suggest that this pattern also explains an MLB career with the initial downturn occurring as individuals are sorted from the league and the upturn due to a decline in ability as players age.

Comparing the three plots shows many crosses. The hazard plot suggests that the hazard rates for black, foreign-born Hispanic, and white players are similar. Interpreting the differences in hazard rates as discrimination, however, is misleading. The problem arises for the same reason that using differences in median wage is misleading in the wage literature: neither method controls for differences due to productivity. In the next section we analyze career duration using semiparametric techniques to control for differences in productivity.

Semiparametric Estimates of Career Duration

We estimate semiparametric hazard functions following Berger and Black (1998), Berger, Black, and Scott (2004), and Groothuis and Hill (2004). Because our data are at the seasonal

Table 2A. Major League Hitter Hazard Rates

Tenure	White	Black	Hispanics
1	0.182	0.115	0.201
2	0.130	0.129	0.142
3	0.109	0.099	0.129
4	0.101	0.070	0.064
5	0.130	0.101	0.071
6	0.070	0.093	0.069
7	0.069	0.102	0.106
8	0.151	0.145	0.040
9	0.174	0.138	0.129
10	0.128	0.125	0.080
11	0.212	0.118	0.087
12	0.276	0.119	0.146
13	0.247	0.231	0.194
14	0.375	0.054	0.045
15	0.303	0.194	0.111
16	0.333	0.207	0.214
17	0.375	0.182	0.250
18	0.308	0.316	0.250
19	0.300	0.182	0
20	0.714	0.286	
21	0.667	0.400	
22	1	0	
23		0	
24		0	
25		0	
26		1	

level, we calculate our hazard model as a discrete random variable. As with Berger, Black, and Scott (2004), we model the durations of a single spell and assume a homogeneous environment so that the length of the spell is uncorrelated with the calendar time in which the spell begins. This assumption lets us treat all the players' tenure as the same, regardless of when it occurred in the panel study. For instance, all fourth year players are considered to have the same baseline hazard regardless of calendar time, so a fourth year player in 1990 has the same baseline hazard as a fourth year player in 1997.

To understand how stock data influence a likelihood function, we follow the notation of Berger, Black, and Scott (2004). Suppose the probability mass function (pmf) of durations is defined as $f(t, x, \beta)$, where t is the duration of the career, x is a vector of performance and personal characteristics, and β is a vector of parameters. Now denote $F(t, x, \beta)$ as the cumulative distribution function; then the probability that a career lasts at least t^o years is simply $1 - F(t^o, x, \beta)$. If we define the hazard function as $h(t, x, \beta) \equiv f(t, x, \beta)/S(t, x, \beta)$, where S is the survivor function $S(t, x, \beta) = \prod_{i=1}^{t-1} [1 - h(i, x, \beta)]$, and apply the definition of conditional probabilities, we may express the pmf as

$$f(t_i, x_i, \beta) = \prod_{j=0}^{t_i-1} [1 - h(j, x_i, \beta)]\, h(t_i, x_i, \beta). \tag{2}$$

If we have a sample of n observations, $\{t_1, t_2, \ldots, t_n\}$, the likelihood function of the sample is

Table 2B. Major League Pitcher Hazard Rates

Tenure	White	Black	Hispanics
1	0.305	0.283	0.236
2	0.241	0.261	0.246
3	0.247	0.152	0.219
4	0.218	0.357	0.185
5	0.190	0.050	0.168
6	0.236	0.182	0.257
7	0.176	0.100	0.158
8	0.204	0.316	0.245
9	0.168	0.214	0.350
10	0.281	0.250	0.357
11	0.185	0.200	0.263
12	0.326	0.125	0.357
13	0.323	0.143	0.400
14	0.373	0.167	0.429
15	0.372	0.200	0.250
16	0.400	0.500	0.667
17	0.174	0.500	0
18	0.529	1	0
19	0.200		0
20	0.444		
21	0.400		
22	0.333		
23	0.333		
24	0.500		
25	0		
26	1		

$$L(\beta) = \prod_{i=1}^{n} f(t_i, x_i, \beta) = \prod_{i=1}^{n} \left(\prod_{j=1}^{t_i - 1} [1 - h(j, x_i, \beta)] h(t_i, x_i, \beta) \right). \tag{3}$$

Often it is not possible to observe all careers until they end, hence careers are often right-censored. Let set A be the set of all observations where the players' careers are completed, and let set B be the set of all observations where the careers are right-censored. For the set of right-censored observations, all we know is that the actual length of the career is greater than t_i, the observed length of the career up through the last year. Because we know that the actual length of the career is longer than we observe, then the contribution of these observations to the likelihood function is just the survivor function (S).

To introduce stock sampling, let set C be the set of careers that were in progress when data collection began. For these observations, we know that career i has lasted for r years before the panel begins so the likelihood must be adjusted by the conditional probability of the career having length r. Of course, some stock-sampled observations may be right-censored. Let set D be the set of all stock-sampled observations that are also right-censored. An example of a career that is both right- and left-censored would be a player that starts his career prior to 1990 and ends his career after 2004. Taking into account all four sets—A, B, C, and D—the likelihood function becomes

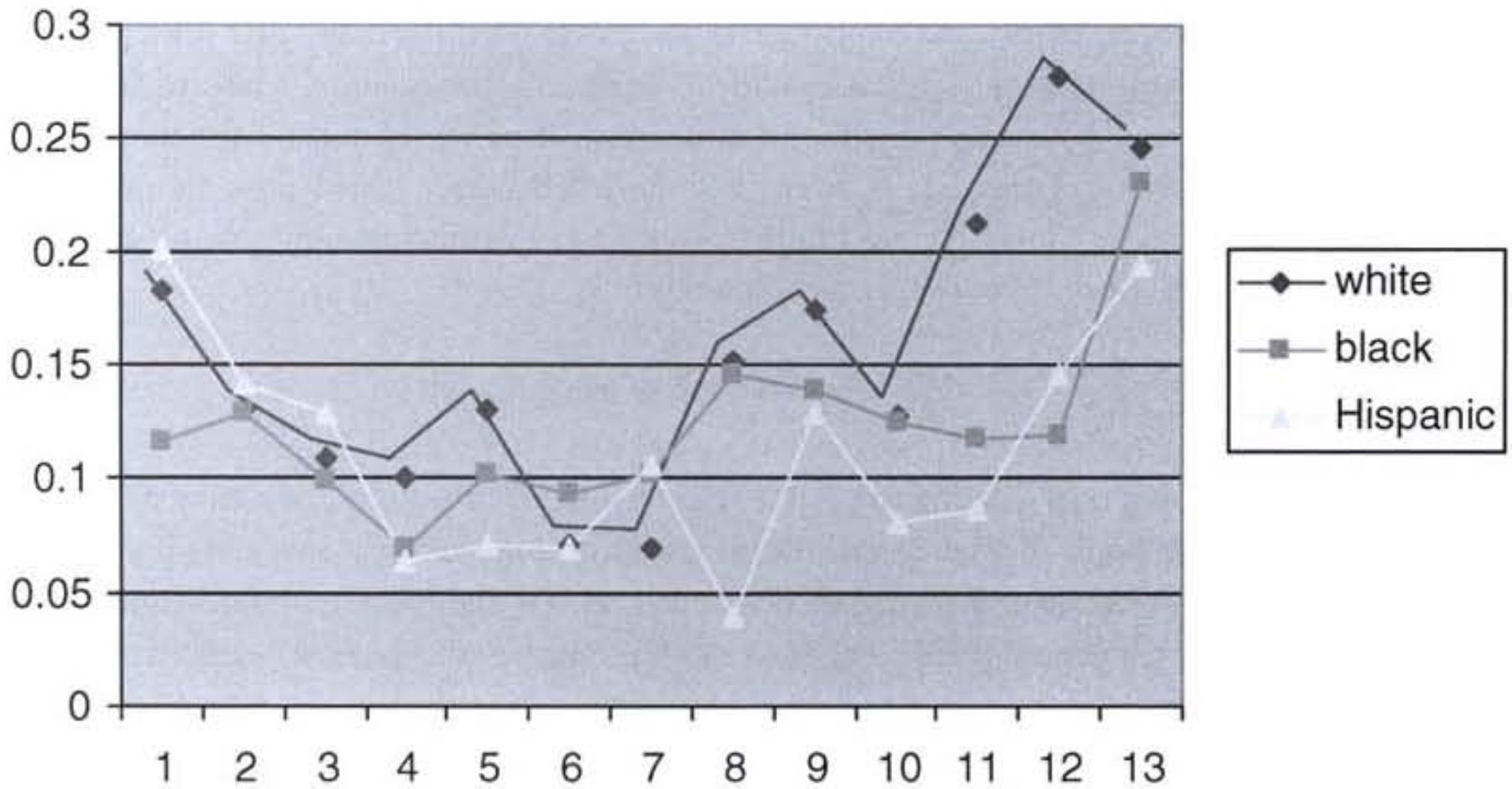

Figure 1A. Hazard Rates for Major League Hitters

$$L(\beta) = \prod_{i \in A} \left(\prod_{j=1}^{t_i - 1} [1 - h(j, x_i, \beta)] h(t_i, x_i, \beta) \right) \times \prod_{i \in B} \left(\prod_{j=1}^{t_i - 1} [1 - h(j, x_i, \beta)] \right)$$
$$\times \prod_{i \in C} \left(\prod_{j=r_i}^{t_i - 1} [1 - h(j, x_i, \beta)] \right) h(t_i, x_i, \beta) \times \prod_{i \in D} \left(\prod_{j=r_i}^{t_i - 1} [1 - h(j, x_i, \beta)] \right). \tag{4}$$

In Equation 4, the contribution of censored, stock-sampled observations to the likelihood function is strictly from the last two terms; such observations simply provide information about the survivor function between (r, t).

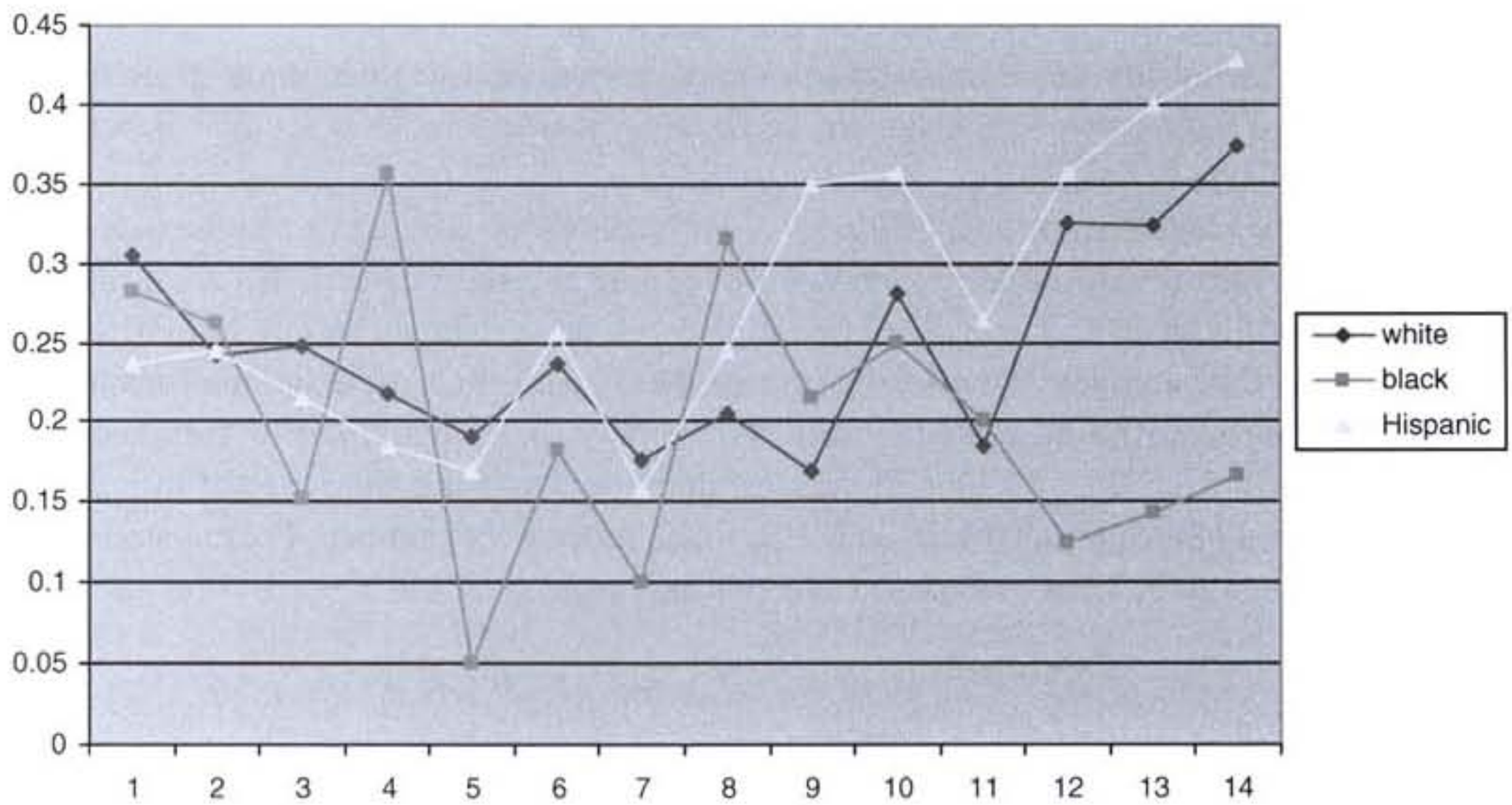

Figure 1B. Hazard Rates for Major League Pitchers

Thus we, like Berger, Black, and Scott (2004), have expressed the likelihood function as a function of the hazard functions. All that remains is to specify the form of a hazard function and estimate by means of maximum likelihood estimation. Because the hazard function is the conditional probability of exiting MLB given that the MLB career lasted until the previous season, the hazard function must have a range from 0 to 1. In principle, any mapping with a range from 0 to 1 will work. Cox (1972) recommends

$$\frac{h(t, x, \beta)}{1 - h(t, x, \beta)} = \frac{h_t}{1 - h_t} e^{x\beta} = \exp(\gamma_t + x\beta), \tag{5}$$

which is simply the logit model with intercepts that differ by period. The term h_t is a baseline hazard function, which is common to all. The $x\beta$ term, determined by the player's personal and productivity characteristics, shifts the baseline hazard function, but it affects the baseline hazard function in exactly the same way each period. Berger and Black (1998) consider other hazard functions and find that the results are relatively robust across various specifications of the hazard function. Because the logit model is available in many software packages, we follow Cox and use the logit model.

The intuition behind Equation 5, when using the logit model for the hazard function, is relatively simple. For each year during the survey in which the player is in MLB, the player either comes back for another season or ends his career. If the career ends, the dependent variable takes on a value of 1; otherwise, the dependent variable is 0. The player remains in the panel until the player exits MLB or the panel ends. If the panel ends, we say the worker's spell is right-censored. Thus, a player who begins his MLB career during the panel and plays for six years will enter the data set six times: The value of his dependent variable will be 0 for the first five years (tenure one through five) and be equal to 1 for the sixth year.

To illustrate a stock sample, consider another player who enters the panel with seven years of MLB job tenure prior to 1990, the first year of the panel, then plays for an additional three years for a 10-year career. For this player, we ignore his first seven years of tenure because he is left-censored. As the equation of the likelihood function with stock data indicates, the duration of a MLB career prior to the beginning of the panel makes no contribution to the value of the likelihood function. Therefore, only years 8 through 10 will enter the data set with the dependent variable taking on the value 0 for years eight and nine and in the tenth year it takes on a value of 1 with this player appearing in the data set a total of three times. Note for all players who are right-censored, we do not know when their career ends so their dependent variables are always coded as 0.

Because the players in the panel have varying degrees of job tenure prior to the beginning of the panel, we identify the hazard function for both long and short careers. The disadvantage to this approach is that the vector γ_t of Equation 5 can be very large. In our study it would require 25 dummy variables. We also run into problems with the Cox technique because we have too few players who have long careers. To simplify the computation of the likelihood function and be able to keep the long careers, we simply approximate the γ_t vector with a fifth order polynomial of the players' tenure in MLB, which reduces the number of parameters to be estimated from 25 to 5. Thus, the hazard function becomes

$$\frac{h(t, x, \beta)}{1 - h(t, x, \beta)} = \Phi(t)e^{x\beta} = \exp[\phi(t) + x\beta], \tag{6}$$

where $\phi(t)$ is a fifth order polynomial of the player's tenure in MLB. The fifth order polynomial therefore includes tenure to the first, second, third, fourth, and fifth powers. Once again, we

Table 3A. Hitter Career Duration Semiparametric Analysis (1990–2004)

Variable	Multinomial Logit Model
Constant	−8.900 (3.87)
Black	−0.191 (1.65)
Foreign-born Hispanic	0.054 (0.57)
Asian-born players	−0.967 (0.84)
Age	0.603 (3.86)
Age squared	−0.007 (2.72)
Slugging average	−2.335 (3.61)
Stolen base percentage	−0.038 (3.14)
Home runs per game	−1.485 (4.50)
Runs batted in per game	−1.158 (2.21)
Walks per game	−0.243 (0.61)
Strikeouts per game	0.889 (3.86)
Hit by pitch per game	−1.243 (1.58)
Sacrifice flies per game	1.106 (0.50)
Runs scored per game	−1.492 (2.63)
Games	−0.018 (11.77)
Infield	−0.374 (3.69)
First base	−0.113 (0.74)
Catcher	−0.902 (6.78)
Primary DH	0.247 (0.92)
χ^2	1611.77
Number observations	7210

First through fifth order tenure polynomials are also included to provide for general functional form of baseline hazard; they are jointly significant. The numbers in parentheses are absolute value of *t*-ratios.

choose the Taylor series approximation technique over using tenure dummies because of the small number of observations for high tenures. This method provides a very flexible specification of the baseline hazard but does impose more restrictions than Cox's model.[3]

Empirical Results for Hitters

In Table 3A, we report the estimates for hitters for Equation 6; in Table 4A, we report the estimates for hitters for Equation 6 by race. In Table 5A, we report the estimates for hitters by race broken down into smaller time increments.

Performance Statistics

In our model of exit, we suggest that improvements in performance should increase career length; owners will keep players who are more productive. The results from Table 3A firmly support this hypothesis. An increase in slugging average, home runs per game, runs batted in per game, runs scored per game, stolen base percentage, and games played all lead to a decrease in the probability of exit. Coefficients for all these variables are properly signed and significant at the 5% or 1% level. Obviously there is a high degree of correlation between home runs, runs batted in, runs scored, and slugging average; however, each variable captures a slightly different

[3] When higher order polynomials of the sixth and seventh power are included, the results do not change, suggesting that a fifth order polynomial is flexible enough to capture the influence of the baseline hazard.

Table 3B. Pitcher Career Duration Semiparametric Analysis (1990–2004)

Variable	Multinomial Logit Model
Constant	−8.74 (4.37)
Black	0.059 (0.32)
Foreign-born Hispanic	−0.213 (1.86)
Asian-born players	−0.227 (0.64)
Age	0.446 (3.24)
Age squared	−0.004 (2.03)
Earned run average	0.121 (6.50)
Wins	−0.147 (5.60)
Losses	0.077 (3.36)
Saves	−0.048 (4.15)
Strikeout-to-walk ratio	−0.350 (6.82)
Outs in innings pitched	−0.0037 (4.50)
Games	−0.018 (6.82)
Left-handed	−0.173 (2.15)
χ^2	1459.50
Number observations	6956

First through fifth order tenure polynomials are also included to provide for general functional form of baseline hazard; they are jointly significant. The numbers in parentheses are absolute value of *t*-ratios.

aspect of offensive performance. It is interesting to note that home runs per game has the highest *t*-value of these variables; this is not surprising given the media attention focused on home runs. An increase in strikeouts per game leads to an increase in exit; the coefficient of this variable is also significant at the 1% level. The coefficients of walks per game and hit by pitch per game are properly signed but not significant; the coefficient of sacrifice flies per game is improperly signed but not significant.

Table 4A. Hitter Career Duration Semiparametric Analysis by Race (1990–2004)

Variable	Logit Model White	Logit Model Black	Logit Model Foreign-Born Hispanic
Constant	−11.61 (3.27)	−9.89 (5.88)	−10.68 (2.32)
Age	0.774 (3.26)	0.575 (1.45)	0.771 (2.37)
Age squared	−0.010 (2.46)	−0.006 (1.01)	−0.010 (1.79)
Slugging average	−2.853 (3.27)	−0.887 (0.65)	−3.430 (2.17)
Stolen base percentage	−0.035 (1.74)	−0.054 (2.63)	−0.036 (0.82)
Home runs per game	−1.632 (3.70)	−0.969 (00.94)	−1.46 (1.92)
Runs batted in per game	−0.142 (0.20)	−3.78 (3.23)	−1.01 (0.82)
Walks per game	−0.506 (0.94)	0.039 (0.04)	0.073 (0.07)
Strikeouts per game	0.587 (1.92)	1.439 (2.62)	0.404 (0.74)
Hit by pitch per game	−0.141 (0.07)	−1.87 (0.44)	−0.631 (0.17)
Sacrifice flies per game	0.486 (0.17)	4.33 (0.84)	−0.680 (0.13)
Runs scored per game	−1.21 (1.60)	−2.54 (1.94)	−0.609 (0.46)
Games	−0.019 (9.67)	−0.016 (4.40)	−0.021 (5.59)
Infield	−0.504 (3.67)	−0.181 (0.75)	−0.449 (1.94)
First base	−0.104 (0.70)	−0.563 (1.49)	−1.10 (0.78)
Catcher	−1.04 (6.30)	−0.583 (0.93)	−1.10 (3.35)
Primary DH	0.754 (2.11)	−0.79 (0.14)	−0.224 (0.28)
χ^2	894.84	409.38	308.02
Number of observations	3795	1621	1462

First through fifth order tenure polynomials are also included to provide for general functional form of baseline hazard; they are jointly significant. The numbers in parentheses are absolute value of *t*-ratios

Table 4B. Pitcher Career Duration Semiparametric Analysis by Race (1990–2004)

Variable	Logit Model White	Logit Model Black	Logit Model Foreign-Born Hispanic
Constant	−7.82 (3.39)	−13.15 (0.82)	−10.59 (2.21)
Age	0.375 (2.37)	0.700 (0.60)	0.588 (1.74)
Age squared	−0.003 (1.33)	0.004 (0.22)	0.008 (1.35)
Earned run average	0.135 (6.28)	0.064 (0.56)	0.083 (1.90)
Wins	−0.157 (5.31)	−0.274 (1.96)	−0.079 (1.01)
Losses	0.080 (3.12)	0.134 (1.07)	0.082 (1.25)
Saves	−0.048 (3.60)	−0.028 (0.72)	−0.078 (1.82)
Strikeout-to-walk ratio	−0.336 (5.96)	0.061 (0.18)	−0.557 (3.33)
Outs in innings pitched	−0.004 (4.16)	−0.002 (0.61)	−0.005 (1.84)
Games	−0.016 (5.45)	−0.029 (2.00)	−0.016 (2.05)
Left-handed	−0.156 (1.76)	−0.339 (0.68)	−0.214 (0.82)
χ^2	1174.32	83.62	108.51
Number observations	5389	294	1008

First through fifth order tenure polynomials are also included to provide for general functional form of baseline hazard; they are jointly significant. The numbers in parentheses are absolute value of *t*-ratios.

Race

In our model of exit, we also suggest that race may play a role if white customers have a preference to see white players and if teams find it difficult to replace white players. Only the coefficient for foreign-born Hispanic players is positive, meaning this attribute increases the probability of exit for the period, but the coefficient is not significant. The coefficients for the

Table 5A. Hitter Career Duration Semiparametric Analysis by Period

Variable	Logit Model 1990–1995	Logit Model 1995–2000	Logit Model 2000–2004
Constant	−13.040 (2.79)	−5.70 (1.43)	−3.18 (0.90)
Black	−.220 (1.17)	−0.098 (0.54)	−0.222 (1.08)
Foreign-born Hispanic	0.291 (1.56)	0.041 (0.24)	−0.044 (0.26)
Foreign-born Asian[a]	—	—	−0.711 (0.63)
Age	0.883 (2.98)	0.342 (1.24)	0.216 (0.91)
Age squared	−0.010 (2.06)	−0.002 (0.45)	−0.002 (0.45)
Slugging average	−2.96 (2.65)	−1.79 (1.77)	−1.69 (1.65)
Stolen base percentage	−0.062 (2.78)	−0.068 (3.11)	−0.011 (0.56)
Home runs per game	−1.215 (2.19)	−1.698 (3.27)	−1.654 (2.98)
Runs batted in per game	−1.425 (1.57)	−1.022 (1.29)	−0.714 (0.82)
Walks per game	−0.499 (0.75)	−0.099 (0.03)	0.355 (0.52)
Strikeouts per game	0.787 (1.98)	1.01 (2.79)	0.978 (32.53)
Hit by pitch per game	−1.77 (0.54)	−2.58 (1.04)	−1.07 (0.45)
Sacrifice flies per game	0.105 (0.03)	2.36 (0.69)	−8.74 (0.23)
Runs scored per game	−2.48 (2.57)	−1.11 (1.22)	−1.56 (1.66)
Games	−0.019 (6.95)	−0.019 (8.05)	−0.015 (6.52)
Infield	−0.593 (3.39)	−0.277 (1.75)	−0.190 (1.16)
First base	0.382 (1.56)	−0.335 (1.31)	−0.256 (1.04)
Catcher	−1.14 (5.17)	−0.894 (4.18)	−0.675 (3.06)
Primary DH	0.273 (0.61)	0.115 (0.26)	0.369 (0.84)
χ^2	755.68	705.63	446.03
Number observations	2727	2937	2533

First through fifth order tenure polynomials are also included to provide for general functional form of baseline hazard; they are jointly significant. The numbers in parentheses are absolute value of *t*-ratios

[a] No foreign-born Asians were in the league in the 1990–1995 and 1995–2000 periods.

Table 5B. Pitcher Career Duration Semiparametric Analysis by Period

Variable	Logit Model 1990–1995	Logit Model 1995–2000	Logit Model 2000–2004
Constant	−10.41 (2.69)	−10.21 (3.21)	−6.18 (2.25)
Black	−0.042 (0.13)	0.176 (0.65)	0.305 (1.08)
Foreign-born Hispanic	−0.106 (0.46)	−0.169 (0.96)	−0.210 (1.37)
Foreign-born Asian[a]	—	—	−0.281 (0.75)
Age	0.552 (2.07)	0.451 (2.08)	0.89 (1.52)
Age squared	−0.006 (1.26)	−0.004 (1.24)	−0.002 (0.79)
Earned run average	0.179 (5.24)	0.167 (5.59)	0.078 (3.02)
Wins	−0.186 (4.20)	−0.074 (1.78)	−0.147 (3.76)
Losses	−0.009 (0.230)	0.111 (3.04)	0.125 (3.69)
Saves	−0.096 (3.96)	−0.067 (3.63)	−0.021 (1.43)
Strikeout-to-walk ratio	−0.609 (5.78)	−0.317 (4.00)	−0.219 (3.28)
Outs in innings pitched	−0.002 (1.57)	−0.007 (5.21)	−0.004 (3.29)
Games	−0.004 (2.09)	−0.007 (2.02)	−0.028 (7.06)
Left-handed	−0.507 (3.67)	−0.045 (0.35)	−0.028 (0.23)
χ^2	596.12	684.69	624.51
Number of observations	2483	2881	3069

First through fifth order tenure polynomials are also included to provide for general functional form of baseline hazard; they are jointly significant. The numbers in parentheses are absolute value of t-ratios.
[a] No foreign-born Asians in our panel exited MLB in the 1990–1995 and 1995–2000 periods.

other race dummy variables are negative, indicating members of these groups have lower probability of exit, but both coefficients are all insignificant.

To convert the coefficient into a percentage and focus on the magnitude of the effect, we use $100[\exp(\beta) - 1]$. This conversion gives us the percentage difference in hazard rates between whites and blacks, between whites and foreign-born Hispanics, and between whites and foreign-born Asians. We find that black hitters have a 17% lower hazard rate than white hitters, while foreign-born Hispanic hitters have a 5.5% higher hazard rate than whites. Asian-born hitters, a very small percentage of the sample, have a 62% lower hazard rate than white hitters. Again, none of the results is statistically significant.

To examine the role of race and duration more fully, we estimated the model separately for white, black, and foreign-born Hispanic players; the sample size of Asian-born players was too small to allow for separate estimation. These estimations, shown in Table 4A, indicate similar results for the three race models. Unfortunately the interaction of correlation between variables and decreased sample size has caused some offensive variables to become insignificant. These results do not indicate any differential treatment between the races based on performance statistics in general over the period.[4]

Estimations of the overall hitter model by smaller time spans are shown in Table 5A. The coefficient of foreign-born Hispanic is positive but not significant in the first period (1990–1995), positive and insignificant in the second period (1995–2000), and negative and insignificant in the last period (2000–2004). In the first period only 16% of the sample is foreign-born Hispanic; this rises to 19% in the second period and 25% in the third period. The coefficient for black players is negative and insignificant in all three periods; the coefficient for foreign-born Asian hitters is negative and insignificant in the only period run.

[4] The race-specific regressions were also run using only slugging average to capture offensive performance. In these models the sign and magnitude of slugging average are very similar for each race. These results are available on request from the authors.

Tenure and Age

In Table 3A, we find that the tenure polynomials, not shown, are jointly significant. Exit probability rises significantly with player age; however, the coefficient of age squared is negative and significant. The signs of these variables do not change in the smaller sample size regressions in Table 4 (by race) and Table 5 (by smaller periods), but the coefficients are not always significant.

Position

Results in Table 3A indicate both infielders and catchers enjoy longer playing careers, *ceteris paribus*, than outfielders (excluded category) and first basemen. This is not surprising because these positions are viewed as more defensive in nature; whereas, first basemen and outfielders are generally viewed as more offensive positions and offensive contributions are already captured by the performance statistics included in the model. In the regression results of the models run for the separate races, shown in Table 4A, the coefficients for infield and catcher are again negative and significant for white and foreign-born Hispanics; the coefficients are properly signed but insignificant in the black model. The vast majority of black hitters are outfielders; only 20.4% of the black sample were infielders. The use of the designated hitter (DH) in the American League is often said to extend the career of aging sluggers who are no longer able to perform adequately in the field. The coefficient of primary DH is not significant in the overall model shown in Table 3A nor in the models run for shorter periods shown in Table 5A; in the regressions broken down by race, shown in Table 4A, the coefficient of primary DH is positive and significant in the white model only; the coefficients of primary DH are negative but insignificant in the black and foreign-born Hispanic models. Data from positional charts of American League teams for each season indicated that 40% of the DH positions were occupied by white hitters (58% of the hitters in the overall sample are white), 33% by black hitters (20% of the hitters in the overall sample are black), and 20% by foreign-born Hispanic hitters (22% of the overall hitter sample are foreign-born Hispanics). In many cases the same player occupied the DH position for multiple seasons. The percentage of black DHs is greater than the percentage of black hitters in our sample; however, the black hitters have better offensive performance statistics on average than white hitters, so this is to be expected.

Comparison to Jiobu's Results

Jiobu found evidence of exit discrimination against black hitters but not Hispanic hitters. We find no evidence of exit discrimination against black hitters, but there is evidence of exit discrimination against foreign-born Hispanic hitters in the early years (1990–2000) of our sample that disappears by the end of our sample period (2000–2004). Several explanations for the difference in results can be posited. First, our data cover a more recent period. Perhaps competitive market forces have removed the last vestige of discrimination against black hitters from MLB. Second, while our econometric technique is similar in nature to that of Jiobu, he used only career performance statistics. None of his data varied between seasons. Therefore, his finding of exit discrimination rests solely on the higher average career performance statistics of black hitters versus white and Hispanic hitters and the average duration of each race. Jiobu finds longer career length on average for black hitters but finds a positive and significant coefficient for

his black dummy variable in the duration analysis. This would seem a foregone conclusion given their higher average career statistics and longer career lengths. Our model has yearly performance statistics that allow the duration analysis to adjust for the "decay rate" of hitter skills.

Empirical Results for Pitchers

In Table 3B, we report the estimates for pitchers for Equation 6; in Table 4B, we report the estimates for pitchers for Equation 6 by race. In Table 5B, we report the estimates for pitchers by race broken down into smaller time increments.

Performance Statistics

The coefficients of all of the performance measures in Table 3B are properly signed and significant determinants of career duration. Increases in earned run average and losses increase the probability of exit while increases in wins, saves, strikeout-to-walk ratio, outs in innings pitched, and games played decrease the probability of exit. All coefficients are significant at the 1% level.

Race

Results for the race coefficients are perplexing. The coefficient for the black dummy is positive but insignificant; the coefficient for Asian-born players is negative but insignificant; the coefficients for foreign-born Hispanics are negative and significant at the 10% level. This statistical anomaly is hard to explain; Hispanic pitchers have longer career lengths than warranted by their performance statistics.

Once again, to convert the coefficient into a percentage and focus on the magnitude of the effect, we use $100[\exp(\beta) - 1]$. This conversion gives us the percentage difference in hazard rates between whites and black. Here we find that black pitchers have a 6% higher hazard rate than whites; although, it is insignificant. On the other hand, foreign-born Hispanic pitchers have lower hazard rates than white pitchers. Foreign-born Hispanic pitchers have a 19% lower hazard rate; the difference is statistically significant at the 5% level. Asian-born pitchers, a very small percentage of the sample, have a 20% lower hazard rate than white pitchers, but this difference is not statistically significant.

As was done for hitters, we estimated the model separately for white, black, and foreign-born Hispanic pitchers; the sample size of Asian-born pitchers was too small to allow for separate estimation. These regressions, shown in Table 4B, indicate similar results for the three race categories. The signs of all pitching variable coefficients are the same in the three race models except for strikeout-to-walk ratio in the black model. It is positive instead of negative but insignificant. Some of the coefficients of the performance variables in the black and foreign-born Hispanic specifications of the model are insignificant, possibly due to the smaller sample size. Coefficients for the other variables (age, age squared, and left-handed) are properly signed but not significant in many cases. These results do not indicate any differential treatment between the races in general over the period.

Estimations of the overall pitcher model by smaller time spans, shown in Table 5A, fail to confirm the evidence of exit discrimination in favor of Hispanic pitchers in any of the smaller time spans. The coefficients are all negative, but none is significant. The coefficient for black pitchers is also negative for the first period but not significant. The coefficient for black pitchers

is positive in the latter two periods but not significant. Given the lack of robustness in these findings, it is difficult to conclude that Hispanic pitchers were given favorable treatment.

Tenure and Age

Exit probability rises with player age but decreases with age squared. The signs of the coefficients of these variables remain the same in all regressions but lack significance in the smaller sample size variations. The tenure polynomials, not shown in the table, are jointly significant again in all specifications of the model.

Left-Handed

Left-handed pitchers appear to enjoy longer careers, *ceteris paribus*, than right-handers. The sign of the coefficient of left-handed is negative in all specifications of the model and is significant at the 5% level in the general model (Table 3B), at the 1% level in the 1990–1995 period (Table 5B), and at the 10% level in the white model (Table 4B); it is not significant in other specifications of the model. Given the small percentage of the population that is left-handed and the lower batting average of left-handed batters against left-handed pitchers, it has been noted by baseball pundits that left-handed pitchers do not have to be as good as right-handers to earn a roster spot in MLB.

5. Conclusion

Overall in our semiparametric duration analysis, we find that performance variables are important in determining career length. Our results for pitchers do not indicate any exit discrimination against the minority groups. We find the anomalous result that foreign-born Hispanic pitchers have longer career lengths than warranted by their performance statistics. This finding is somewhat tenuous at best; the coefficient is only significant at the 10% level in the overall sample and does not appear at any significant levels in the smaller time span breakdowns.

Past research suggested that discrimination by majority white fans led owners in sports to keep less talented white players on rosters. Our findings are inconsistent with results from a study by Jiobu (1988) in a prominent sociology journal (*Social Forces*) using 1971–1985 data that found that race decreased career length, *ceteris paribus*, for black hitters but not foreign-born Hispanics. We find no evidence that race affects the career duration of black or foreign-born Hispanic hitters. Our methodology allows the duration analysis to adjust for the "decay rate" of hitter skills. It is impossible to determine if the difference in our findings is the result of our newer data set or use of yearly rather than career performance statistics. Our results could be interpreted as an affirmation of Becker's theoretical implications of market competition overcoming discrimination.

References

Becker, Gary. 1971. *The economics of discrimination.* 2nd edition. Chicago: University of Chicago Press.

Berger, Mark C., and Dan A. Black. 1998. The duration of Medicaid spells: An analysis using flow and stock samples. *The Review of Economics and Statistics* 80:667–75.

Berger, Mark, Dan Black, and Frank Scott. 2004. Is there job lock? *Southern Economic Journal* 70:953–76.

Bodvarsson, Orn B., and Ramond T. Brastow. 1999. A test of employer discrimination in the NBA. *Contemporary Economic Policy* 17:243–55.

Bolla, G. A. 2002. Accelerated useful life testing and field correction methods. Visteon Corporation Paper, Society of Automotive Engineers.

Cox, D. R. 1972. Regression models with life tables. *Journal of Royal Statistical Society* 34:187–220.

Cymrot, Donald J. 1985. Does competition lessen discrimination? Some evidence. *Journal of Human Resources* 20:605–12.

Groothuis, Peter A., and J. Richard Hill. 2004. Exit discrimination in the NBA: A duration analysis of career length using flow and stock samples. *Economic Inquiry* 42:341–9.

Hill, James Richard. 2004. Pay discrimination in the NBA revisited. *Quarterly Journal of Business and Economics* 43:81–92.

Hill, James Richard, and William Spellman. 1984. Pay discrimination in baseball: Data from the seventies. *Industrial Relations* 23:103–12.

Hoang, Ha, and Dan Rascher. 1999. The NBA, exit discrimination, and career earnings. *Industrial Relations* 38:69–91.

Jiobu, Robert M. 1988. Racial inequality in a public arena: The case of professional baseball. *Social Forces* 67:524–34.

Johnson, Bruce K. 1992. Team racial composition and players' salaries. In *Diamonds are forever: The business of baseball*, edited by Paul M. Sommers. Washington, DC: The Brookings Institute, pp. 189–202.

Johnson, Norris R., and David P. Marple. 1973. Racial discrimination in professional basketball: An empirical test. *Sociological Focus* 6:6–18.

Kahn, Lawrence M. 1991. Discrimination in professional sports: A survey of the literature. *Industrial and Labor Relations Review* 44:395–418.

Kahn, Lawrence M., and Peter D. Sherer. 1988. Racial differences in professional basketball players compensation. *Journal of Labor Economics* 6:40–61.

Kanazawa, Mark T., and Jonas P. Funk. 2001. Racial discrimination in professional basketball: Evidence from the Nielsen ratings. *Economic Inquiry* 39:599–608.

Nardinelli, Clark, and Curtis Simon. 1990. Customer discrimination in the market for memorabilia: The case of baseball. *Quarterly Journal of Economics* 105:575–95.

Oaxaca, Ronald. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* 14:693–709.