IZA

DISCUSSION PAPER SERIES

IZA DP No. 5989

Cross-Assignment Discrimination in Pay: A Test Case of Major League Baseball

Örn B. Bodvarsson
John G. Sessions

September 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Cross-Assignment Discrimination in Pay: A Test Case of Major League Baseball

Örn B. Bodvarsson
St. Cloud State University and IZA

John G. Sessions
University of Bath and IZA

Discussion Paper No. 5989
September 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5989
September 2011

ABSTRACT

Cross-Assignment Discrimination in Pay: A Test Case of Major League Baseball*

The traditional Becker/Arrow style model of discrimination depicts majority and minority and workers as perfectly substitutable inputs, implying that all workers have the same job assignment. The model is only appropriate for determining whether pay differences between, for example, whites and non-whites doing job assignment A are attributable to prejudice ('within-assignment discrimination'); It is inappropriate, however, for determining whether pay differences between whites in job assignment A and non-whites in job assignment B reflect discriminatory behaviour ('cross-assignment discrimination'). We test the model of such cross assignment discrimination developed by Bodvarsson and Sessions (2011) using data on Major League Baseball hitters and pitchers for four different seasons during the 1990s, a decade during which monopsony power fell. We find strong evidence of *ceteris paribus* racial pay differences between hitters and pitchers, as well as evidence that cross-assignment discrimination varies with labour market structure.

NON-TECHNICAL SUMMARY

The standard approach to measuring pay discrimination is to test a model of majority/minority pay differences that assumes all workers are perfect substitutes, i.e. have precisely the same human capital endowments and perform precisely the same task. The standard empirical approach then is to compare groups of minority and majority workers within the same occupation or job classification, e.g. comparing white and non-white nurses or male and female airline pilots. The standard model and test are thus unsuitable for assessing discrimination across occupations and job classifications, e.g. testing for whether the reason white physicians make much higher income than non-white nurses is because of gender discrimination. We call this type of discrimination "cross-assignment discrimination" and in a previously published paper, we develop the theory behind this approach to studying discrimination. This paper provides a test of that theory and our chosen test case is Major League Baseball, where white (non-white) hitters (pitchers) are compared with non-white (white) pitchers (hitters). We find strong evidence of cross-assignment discrimination between hitters and pitchers.

JEL Classification: J7

Keywords: wage discrimination, complementarity, monopsony power

Corresponding author:

Örn B. Bodvarsson
Department of Economics
Department of Management
St. Cloud State University
St. Cloud
Minnesota 56301-4498
USA
E-mail: obbodvarsson@stcloudstate.edu

* We thank William Boal, Sabien Dobbelaere, Lawrence Kahn, Wing Suen, Ted To and session participants at the 2007 SOLE and EALE meetings, the 12th IZA/CEPR European Summer Symposium in Labour Economics (ESSLE) and the 2011 RES Annual Conference, for helpful comments on earlier drafts of this paper. We also thank Bree Dority O'Callaghan and Robert Girtz for assistance with data collection. The normal disclaimer applies.

1. Introduction

According to the U.S. Bureau of Labour Statistics, the median weekly earnings of male and female elementary and middle school teachers in 2006 were $920 and $824, respectively, whereas for male and female school principals and school district superintendents, median weekly earnings were $1275 and $1107, respectively. During the same year, the median weekly earnings of male and female registered nurses were $1074 and $971, respectively, whereas for male and female physicians and surgeons they were $1847 and $1329, respectively. Median weekly earnings of male and female lawyers were $1891 and $1333, respectively, whereas for female legal assistants, they were $726 (data on earnings of male legal assistants are not available). Median weekly earnings of male and female cooks were $377 and $340, respectively, whereas for male and female restaurant waitpersons they were $284 and $348, respectively. Finally, during 2006 the median weekly earnings of male aircraft pilots and flight engineers were $1419, whereas for female flight attendants, median weekly earnings were $488 (data on male flight attendant earnings are not available).[1]

What do the above examples have in common? First, each example involves a pair of job assignments within a firm that are distinctly *complementary*; Pilots and flight attendants are complementary labour inputs in the production of airline services, educational administrators and teachers are complements in the provision of educational services whilst physicians and nurses complement one another in the provision of health care services. Second, in each example for which data on earnings of each gender are available, there are noticeable gender pay gaps *within* job assignments – 9 per cent for school teachers, 20 per cent for principals and superintendents, nearly 10 per cent for registered nurses, 28 per cent for physicians and surgeons, nearly 30 per cent for lawyers, 10 per cent for cooks and 22.5 per cent for waitpersons (in favour of females, however). A commonly asked question would

[1] These numbers are taken from the Bureau of Labour Statistics website (http://www.bls.gov/cps/cpsaat39.pdf).

be: How much of these *intra-job* gender pay gaps are attributable to discrimination? This is the approach taken in the traditional wage discrimination model, due originally to Becker (1971) and Arrow (1973). This model is based on the fundamental assumption that majority and minority workers are perfect substitutes in production. Consequently, the traditional model is only appropriate for studying gender, racial, age, sexual orientation or other group pay differences for workers performing the same job assignment.

In this paper, we address a different and more nuanced question: To what extent is majority/minority pay *across* complementary job assignments within a firm attributable to discrimination? For example, how much of the $931 (65.6%) pay gap between male aircraft pilots and flight engineers and female transportation attendants, the $1165 pay gap between male lawyers and female legal assistants and the $876 pay gap between male physicians and surgeons and female registered nurses, attributable to discrimination? Are these gaps primarily attributable to majority/minority productivity differences or to prejudice? This is a question about *inter-job* wage discrimination and it is a far more difficult question because to answer it we need to compare majority and minority workers for which there will be both distinct productivity and labour supply differences. In the traditional (*intra-job* assignment) model of wage discrimination, details of the production function are dispensed with because there are no productivity and labour supply differences between workers. In a study of discrimination across job assignments, however, the production and labour supply functions must be given explicit consideration.

In what follows we empirically test the model of pay discrimination across job assignments developed by Bodvarsson and Sessions (2011) - hereafter BO - on an industry characterized by complementary job assignments, racial integration, variation in monopsony power across worker groups and a history of racial discrimination – U.S. Major League Baseball.

We employ a novel, two-stage regression methodology in which a standardised measure (i.e. common) measure of productivity is estimated separately for each occupation. We then incorporate this measure as a right-hand-side explanatory variable in a second-stage, all-occupation regression designed to estimate cross-assignment discrimination. Our empirical analysis finds convincing evidence of racial differences in pay across player job assignments, even after controlling for a wide array of demographic variables and position-specific productivity. Moreover, we find strong evidence of BO's theoretical prior that racial pay differentials across assignments are affected by changes in relative productivities.

The paper is set out as follows: Section 2 discusses some of the previous literature on the economics of discrimination whilst Section 4 outlines our test case of Major League Baseball. Our empirical analysis is presented in Section 5 whilst final comments are collected in Section 6.

2. Previous Literature

While most of the literature on discrimination has focused on the measurement of the majority/minority pay gap within the same job category, some researchers have suggested that the required assumption of perfect substitution between inputs may be somewhat inappropriate. Indeed, Becker alluded to this issue by sketching a brief extension to his two-factor black/white worker model to a three-factor model [see Becker (1971, pp. 59-62)]. Two of the factors are perfectly substitutable blacks and whites that belong to a group that could be termed 'Type 1 Labour.' Then, there is a third labour input, 'Type 2 Labour,' that both discriminates against blacks and is complementary to, or imperfectly substitutable, for them. Type 2 workers could, for example, be managers. In this situation, Becker showed that there would be a *ceteris paribus* black/white wage gap within the Type 1 category. Arrow (1973) elaborated on this by showing that the black/white wage gap depends upon the sensitivity of

Type 2 labour's reservation wage to the fraction of the firm's labour force that is black, as well as the importance of Type 2 labour as an input (importance is measured as the size of the payments to Type 2 labour relative to Type 1 labour). Neither Becker nor Arrow tested these propositions, nor did they investigate further the implications of complementarity in production for the black/white pay differential.

Welch (1967) raised the possibility that blacks and whites working in the same firm may not be perfect substitutes because there may be differences in their educational endowments. Welch suggested that, perhaps because of long-term discrimination, blacks may have acquired less schooling and/or attended lower quality schools. He modelled educational endowments and physical labour as separate factors of production, allowing for racial differences in educational endowments and, following Becker and Arrow, white co-worker discrimination. He argued that if firms choose racially integrated labour forces then blacks and whites must be complementary inputs. The intuition is that because of whites' aversion to working with blacks, integration creates inefficiencies that will cause joint product to be less than the sum of individual black and white worker marginal products. The firm will therefore follow an apartheid employment policy unless there are sufficiently large complementarities to be exploited, i.e. if the gains from complementarity exceed the losses attributable to co-worker discrimination.[2]

More recently, Kahn (1991) sets out a model of customer discrimination in which whites and blacks are represented as different inputs in the production function. He models blacks and whites as distinct inputs because if customers are prejudiced, they will act as if the amount of black input is equal to just a fraction of the input of *otherwise identical* white workers. Similarly, Bodvarsson and Partridge (2001) present a model of a professional sports

[2] Borjas (2008) also suggested that differential educational attainments may render black and white workers as imperfect substitutes: 'The two groups of workers might have different productivities because they might differ in the amount and quality of educational attainment, or because they might have been employed in different occupations and hence are entering (a) firm with different types of job training. [Borjas (2008), p. 128)].

team where white and non-white athletes are imperfect substitutes due to racial differences in prior training and experience.[3]

An extensive empirical literature on wage discrimination emerged during the 1970s, all based on the original Becker-Arrow model of perfect substitution. The accumulating evidence was called into question in the early 1980s, however, as a number of studies concluded that racial and ethnic groups were not perfectly substitutable. These studies typically applied econometric models of Translog or Generalized Leontief aggregate production functions to estimate elasticities of complementarity between groups. Grant and Hamermesh (1981), for example, found that black adults are imperfect substitutes for white men and complements to white women and youths; Borjas (1983) provided evidence which suggested that whilst black males were imperfect substitutes for white males, Hispanic males and white males were complementary; Borjas (1987) showed that black natives are imperfect substitutes for white natives; and Kahanec (2006) founds that non-whites are complementary to whites.

The traditional empirical approach for testing wage discrimination is generally unsuitable where cross-assignment discrimination is concerned because it is based on a presumption that whites and non-whites are perfect substitutes. While empirical researchers have usually controlled for job assignment differences with dummy variables, that approach has severe limitations because it fails to adequately control for the structure of the underlying production function. As Hashimoto and Kochin (1980) argue, failure to account for any and all sources of productivity differences will lead to biased estimates of discrimination.

In a recent theoretical contribution, BO extend the traditional Becker-Arrow model to ascertain how predictions regarding cross-assignment discrimination vary with the form of

[3] Both of these models, however, have features that limit their applicability. In Kahn's model, whites and non-whites are assigned the same job and would be perfect substitutes if customers were unprejudiced whilst Bodvarsson and Partridge impose the restriction that the cross elasticity of demand for white labour with respect to non-white labour is negative.

the production function. Using an approach similar to Kahn (1991), BO measure the extent of customer prejudice against non-white workers by a parameter, *D*.[4] Customer prejudice may be interpreted as a situation in which customers discount the marginal revenue product (MRP) of non-white workers. The lower (higher) is *D*, the more (less) intense is the prejudice and the lower (higher) is non-white MRP. Prejudice dissipates as *D* approaches 1 and reaches a maximum as *D* falls to 0. While it is traditional to think of customer discrimination as implying a price discount on the output of non-white workers, the approach above is equivalent. The parameter *D* reflects the idea that non-white labour is valued less when customers are prejudiced.[5] In terms of the Generalized Leontief function (GLF), for example, the impact of D is seen as follows:

$$Q = \sum_{i=1}^{k} \sum_{j=1}^{k} g_{ij} [X_i^W \left(D X_j^{NW} \right)]^{\frac{1}{2}} \tag{1}$$

where $D \leq 1$, Q is output, X_i^W is the quantity of white labour input *i*, X_j^{NW} is the quantity of non-white labour input *j*, and g_{ij} is the technology coefficient. Note there are a total of 2*k* inputs – two groups of workers within each job assignment (white and non-white) x *k* job assignments.

[4] Note that prejudice is a necessary but not sufficient condition for *discrimination*. It is only when prejudicial thoughts are acted upon through, for example, exercising product market demand that they can result in discriminatory outcomes in the labour market. In general, taste discrimination in pay and hiring is a market outcome that results from employers acting upon their own racial preferences and/or implementing the racial preferences of customers or co-workers.

[5] BO's approach implies that consumers can discern the racial characteristics of workers when purchasing or consuming the particular good or service in question. Such an assumption is not unrealistic and examples abound of environments in which such an approach to discounting non-white MRP is likely to hold. At professional sports events, white (non-white) fans witness non-white (white) players' contribution to athletic entertainment. If sports fans of one skin colour are prejudiced against players of another colour, this may result in lower pay to the latter group. A similar situation may arise in other entertainment services, e.g. films, theatre, popular music. More generally, there are many production situations in which consumers must interact with minority workers in order for a good or service to be dispensed, e.g. white patients interacting with non-white nurses or doctors, non-white clients interacting with white legal advisers, and white airline passengers interacting with non-white flight attendants. There will also be cases where prejudiced white consumers may not necessarily see non-white workers during the act of purchase or consumption, but mere knowledge of the racial composition of the work force may influence buying decisions. For example, white consumers may place a lower valuation on, or even refuse to purchase, food products, or appliance repair services, or the processing of important financial transactions, knowing that those goods or services were manufactured or performed by non-white workers.

BO then apply Becker's (1971) *Market Discrimination Coefficient* (MDC) to the case of discrimination across job groups. The MDC measures the *ceteris paribus* racial earnings gap viz. the percentage earnings premium paid to whites. If the white and non-white wage is denoted by r^i, $i = W, NW$, then the MDC is given by:

$$MDC_{NW}^{W} = \frac{r^{W}(D<1)}{r^{NW}(D<1)} - \frac{r^{W}(D<1)}{r^{NW}(D<1)} \qquad (2)$$

The first term on the right-hand side of (2) is the wage ratio when there is prejudice (i.e. when D <1) whereas the second term is the wage ratio in the absence of prejudice (i.e. when D = 1). The MDC is the difference between the two ratios and measures the *ceteris paribus* racial pay gap.

Applying equation (2) to the case of cross-assignment discrimination, the *ceteris paribus* racial pay gap between whites performing job 1 and non-whites performing job 2 is:

$$MDC_{NW_2}^{W_1} = \frac{r_1^{W}(D<1)}{r_2^{NW}(D<1)} - \frac{r_1^{W}(D=1)}{r_2^{NW}(D=1)} \qquad (3)$$

Similarly, the *ceteris paribus* racial pay gap between whites performing job 2 and non-whites performing job 1 is:

$$MDC_{NW_1}^{W_2} = \frac{r_2^{W}(D<1)}{r_1^{NW}(D<1)} - \frac{r_2^{W}(D=1)}{r_1^{NW}(D=1)} \qquad (4)$$

BO derive the above measure of cross-assignment discrimination for four different production functions - Generalized Leontief, Quadratic, CES, and Cobb–Douglas. The Generalized Leontief provides the most general results, although closed form solutions are not possible. Closed form solutions, which are obtainable from the other three functions but only under restrictive assumptions, suggest that most predictions are generally robust across functional forms and that cross-assignment discrimination depends upon productivity and labour supply differences between the various worker groups, labour market structure, and the interaction between relative group productivity and prejudice. A uniform prediction across all four production functions is that changes in the relative productivity of one racial group induce changes in cross-assignment discrimination. For example, in all four cases, higher white (non-white) productivity raises (lowers) the amount of discrimination.[6] This is an important finding, both academically and in terms of policy. If non-whites are able to improve their skill-base, or if technological progress impacts more favorably on non-whites relative to whites, then cross-assignment discrimination may be reduced. An increase in white productivity, however, will lead to an unintended adverse consequence by increasing discrimination against non-whites.

Table 1 following summarises BO's various comparative static results for the *ceteris paribus* white/non-white pay differential (i.e. $MDC_{NW_2}^{W_1}$) derived from the four production functions:

Table 1: BO's Comparative Static Results for Cross-Assignment Discrimination $\left(\P MDC_{NW_2}^{W_1}/\P Variable\right)$

Variable	*Generalized Leontief*	*Quadratic*	*CES*	*Cobb-Douglas*
Strength of Prejudice (D)	-	-	-	-
White productivity	+	+	+	±
Non-white productivity	-	-	-	±
White productivity x D	-	-	-	
Non-white productivity x D			+	
White labour supply			-	
Non-white labour supply			+	
White reservation wage				+
Non-white reservation wage				-
Employer's monopsony power				±
Degree of monopsonistic wage discrimination				-

BO's findings have an important general implication: Researchers must control for both

[6] Whilst BO frame their theoretical model in terms of racial discrimination, it is clearly applicable to other types of labour market discrimination, for example, where workers are discriminated against on account of their age, gender, nativity status, sexual orientation, religious affiliations, or other characteristics that may be targets of employer, employee, or consumer prejudice.

productivity differences between white and non-white workers, as well as the interaction between race and productivity, when estimating the extent of cross-assignment discrimination.

2. A Test Case: Major League Baseball

In order to test empirically the BO model of cross-assignment discrimination, we searched for an appropriate test case viz. an industry where: (i) there are accurate data on salaries and productivity for individual workers across distinct job assignments and these data are available for different firms; (ii) the productivities of job assignment groups within the firm are interrelated; (iii) there is racial integration; (iv) the pay of some workers is competitively determined, whilst the pay of others is determined under conditions resembling monopsony; (v) there is potential for customer discrimination; and (vi) there have been changes in the number of employers in the industry over time.

One industry satisfying all these criteria is Major League Baseball (MLB) in the USA.[7] In MLB, each team requires two distinctly complementary types of player skill - hitting (an offensive skill) and pitching (a defensive skill) - in the production of baseball entertainment.[8] Player salaries are set under two different regimes, one competitive, the other monopsonistic. The monopsonistic regime applies to players with fewer than six years of MLB experience. These players are subject to the *reserve clause* and are constrained to negotiate their pay with only one team. The competitive regime applies to players with at least 6 years of MLB experience. They are eligible to file for *free agency* and may negotiate with any team in the league. Monopsony power effectively begins to erode, however, as early

[7] Racial discrimination in professional sports has received considerable attention among labour economists because of the abundant statistical evidence on a player's personal attributes, compensation and productivity. Most studies in this area have focused on discrimination with respect to pay, hiring, retention and positional segregation. For an examination of the research prior to 2000, see Kahn's (2000) expository survey.

[8] Woolway (1997) and Zech (1981) argue that the Cobb-Douglas function is a particularly appropriate description of an MLB team's production situation. They both estimated Cobb-Douglas functions where the dependent variable is team winning percentage and the independent variables are player and team career statistics.

as the fourth year because then a player is eligible for *final offer arbitration*. Arbitration rights tend to relieve players of monopsonistic exploitation because arbitrators strive to award competitive salaries. Pitchers have historically been disproportionately white, whereas the pool of hitters has tended to be more racially balanced. The Major League added new teams (called 'expansion teams') since the early 1990s, leading to a reduction in each team's degree of monopsony power held over reserve clause players.

The ideal way to measure a Major League player's marginal revenue product (MRP) is by his contribution to the team's ticket, broadcasting and merchandise revenues. Because of the team production nature of baseball, however, it is impossible to empirically disentangle one player's revenue contribution from another. We thus proxy MRP by the player's years of MLB experience, tenure with his current team, and various career statistics (computed on a game-by-game basis since the beginning of the player's Major League career) that proxy his ability and skills. The career statistics we use to measure a hitter's productivity include *at bats, stolen bases, bases on balls, total bases, slugging average* and *batting average*. We distinguish between hitters that are 'designated hitters' from those who are not. A designated hitter is a player who is chosen at the start of the game to bat in lieu of the pitcher in the line-up. We also distinguish, using dummies, between hitters that serve other types of positions. These include whether the hitter served as an infielder or a catcher. We measure a pitcher's productivity by use of the following career statistics: *Wins, Losses, Games Started, Complete Games, Saves, Homeruns, Walks, Strikeouts, Innings Pitched, Earned Run Average (ERA,) and Strikeout Rate*. An explanation of baseball terminology is set out in the Appendix.

3 Empirical Analysis

3.1 *Descriptive Statistics*

Tables 2 and 3 present descriptive statistics for hitters and pitchers, respectively. Our full sample comprises 1093 hitters (549 white, 367 black and 177 Hispanic) and 1204 pitchers (942 white, 127 black and 135 Hispanic). Salary, experience, performance and position data were drawn from the *Lahman Baseball Database* (see: www.baseball1.com) over four seasons - 1992, 1993, 1997 and 1998. The Major League expanded by two teams between 1992 and 1993 and again by two teams between 1997 and 1998. The salary data do not include information about contract length, bonus clauses or endorsements. Salaries for players on the Canadian teams were converted to U.S. dollars. The experience data were used to determine the player's eligibility for free agency and final offer arbitration and the player's race was inferred from inspection of *Topps* baseball cards for all four seasons. For the U.S. teams, metropolitan area population and per-capita income were obtained from the website of the Bureau of Economic Analysis (see: www.bea.gov). For the Canadian teams, similar data were obtained from the Statistics Canada website (see: www.statcan.ca). Per-capita income data for the Canadian cities were converted to U.S. dollars.

It would appear from Table 2 that there are no major differences between the personal and professional characteristics of white hitters, black hitters and Hispanic hitters, nor in the characteristics of the greater metropolitan area in which they play. In terms of career characteristics, however, black hitters record significantly more *At Bats*, *Stolen Bases, Bases on Balls* and *Total Basses* than either white hitters or Hispanic hitters. They are also less likely to play as an infielder or catcher, but more likely to play as an outfielder or designated hitter. Compared to Hispanic hitters, white hitters record significantly more *At Bats*, *Bases on Balls* and *Total Bases*, but significantly fewer *Stolen Bases*. They are also more likely to play as a catcher, but less likely to play as an outfielder or designated hitter.

Table 2: Descriptive Statistics: Hitters

Variable	All		White		Black		Hispanic	
	Mean	*Std. Dev*	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*
Personal Characteristics								
Log Annual Salary	13.890	1.13	13.865	1.10	13.938	1.13	13.866	1.22
Age	30.304	3.70	30.596	3.49	30.488	3.95	29.023	3.55
White	0.502	0.500	-	-	-	-	-	-
Black	0.336	0.472	-	-	-	-	-	-
Hispanic	0.162	0.369	-	-	-	-	-	-
Professional Characteristics								
MLB Experience	7.061	3.89	7.062	3.87	7.223	4.07	6.723	3.55
MLB Experience-Squared	64.957	69.31	64.785	70.06	68.684	74.23	57.763	54.59
Tenure with Current Club	2.672	3.00	3.062	3.38	2.305	2.62	2.226	2.24
Free Agent	0.600	0.49	0.598	0.49	0.605	0.49	0.599	0.49
Eligible for Final Offer Arbitration	0.296	0.46	0.304	0.46	0.294	0.46	0.271	0.45
American League	0.514	0.50	0.521	0.50	0.469	0.50	0.588	0.49
National League	0.486	0.50	0.479	0.50	0.057	0.23	0.124	0.33
Canadian Team	0.073	0.26	0.067	0.25	7.223	4.07	6.723	3.55
Performance								
At Bats	2506.414	2001.58	2419.738	1940.51	2699.202	2198.95	2375.525	1720.23
Stolen Bases	69.746	112.52	44.800	72.35	111.055	157.89	61.480	69.63
Bases on Balls	254.275	247.74	253.131	233.32	285.349	293.87	193.39	161.14
Total Bases	1060.200	913.52	1016.772	880.39	1162.845	1013.19	982.073	771.85
Slugging Average	0.407	0.06	0.404	0.06	0.416	0.06	0.397	0.07
Batting Average	0.267	0.03	0.264	0.02	0.271	0.02	0.266	0.02
Infielder	0.459	0.50	0.556	0.50	0.281	0.45	0.531	0.50
Outfielder	0.383	0.49	0.217	0.41	0.657	0.48	0.333	0.47
Catcher	0.116	0.32	0.189	0.39	0.016	0.13	0.096	0.30
Designated Hitter	0.059	0.24	0.046	0.21	0.079	0.27	0.056	0.23
Greater Metro Area Characteristics								
Percentage White	80.507	6.89	80.938	6.77	80.683	6.72	78.808	7.39
Percentage Black	13.273	6.58	12.959	6.60	13.676	6.62	13.409	6.44
Percentage Hispanic	10.621	10.65	10.719	10.80	10.331	10.58	10.918	10.36
Average Annual Income ($)	25562.990	3789.65	25508.570	3757.99	25551.300	3731.59	25756.00	4016.17
Population[1]	5514009	4657988	5313189	4509095	5513759	4729589	6137413	4927354
Year Dummies								
1992	0.250	0.43	0.255	0.44	0.243	0.43	0.249	0.43
1993	0.235	0.42	0.248	0.44	0.237	0.43	0.192	0.40
1997	0.260	0.44	0.248	0.43	0.270	0.44	0.277	0.45
1998	0.255	0.44	0.250	0.43	0.251	0.43	0.282	0.45
Sample Size	1093		549		367		177	

Note:
1. Population denotes the greater metro area population;
2. Source: All variables except Race and Greater Metro Area Characteristics (GMAC) extracted from the Lahman Baseball Database (Version 5.0, Release Date: Dec. 15, 2002). Race is derived form observed Topps Baseball Cards, years 92, 93, 94, 97, 99 (only years available). GMAC derived from the Statistical Abstract 1997-1999, the BEA, CA1-3, and from Statistical Canada..

Table 3: Descriptive Statistics: Pitchers

Variable	All		White		Black		Hispanic	
	Mean	*Std. Dev*	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*	*Mean*	*Std. Dev*
Personal Characteristics								
Log Annual Salary	13.409	1.19	13.451	1.20	13.238	1.16	13.276	1.18
Age	29.815	4.09	30.190	4.02	29.016	4.00	27.948	4.03
White	0.782	0.41	-	-	-	-	-	-
Black	0.105	0.31	-	-	-	-	-	-
Hispanic	0.162	0.37	-	-	-	-	-	-
Professional Characteristics								
MLB Experience	5.988	4.20	6.158	4.20	5.772	4.49	5.000	3.75
MLB Experience-Squared	53.468	76.64	55.562	78.38	53.331	75.31	38.985	63.34
Tenure with Current Club	1.924	2.07	1.935	2.10	1.843	1.97	1.926	1.99
Free Agent	0.467	0.50	0.482	0.50	0.441	0.50	0.385	0.49
Eligible for Final Offer Arbitration	0.306	0.46	0.314	0.46	0.236	0.43	0.319	0.47
American League	0.513	0.50	0.518	0.50	0.543	0.50	0.452	0.50
National League	0.487	0.50	0.475	0.50	0.528	0.50	0.556	0.50
Canadian Team	0.069	0.25	0.063	0.24	0.055	0.23	0.126	0.33
Performance								
Starter	0.442	0.50	0.441	0.50	0.402	0.49	0.489	0.50
Wins	37.446	44.33	39.007	45.27	34.386	42.41	29.430	38.34
Losses	34.179	37.05	35.904	38.37	29.236	30.11	26.785	32.12
Games Started	74.12	105.53	77.769	108.53	59.646	92.16	62.274	93.98
Complete Games	10.15	22.24	10.981	23.33	6.433	14.87	7.844	19.65
Shutouts	2.875	6.08	3.065	6.32	1.984	4.74	2.385	5.35
Saves	19.488	51.87	20.941	52.93	19.362	62.60	9.474	26.16
Homeruns	56.517	62.57	58.842	64.46	50.409	52.94	46.044	56.11
Walks	225.779	249.73	231.782	257.66	224.095	217.58	185.474	217.41
Strikeouts	436.641	514.13	450.726	530.21	436.047	490.18	338.919	402.35
Innings Pitched	627.59	702.43	655.160	720.78	558.969	620.14	499.785	627.21
ERA	4.025	0.96	3.995	0.94	4.175	1.11	4.094	0.97
Strikeout Rate	0.078	0.02	0.078	0.02	0.083	0.02	0.079	0.02
Greater Metro Area Characteristics								
Percentage White	80.714	6.84	80.695	6.91	80.335	6.56	81.201	6.59
Percentage Black	13.038	6.46	12.946	6.49	14.026	6.46	12.750	6.19
Percentage Hispanic	10.975	10.77	10.899	10.61	10.909	10.40	11.573	12.20
Average Annual Income ($)	25488.2	3939.85	25491.51	3895.30	25852.23	3898.44	25122.19	4271.98
Population[1]	5551948	4683875	5481401	4631793	6035905	4915887	5588930	4829139
Year Dummies								
1992	0.221	0.42	0.236	0.42	0.189	0.39	0.148	0.36
1993	0.239	0.43	.248	0.43	0.244	0.43	0.170	0.38
1997	0.264	0.44	.256	0.44	0.276	0.45	0.311	0.46
1998	0.276	0.45	.260	0.44	0.291	0.46	0.370	0.48
Sample Size	1204		942		127		135	

Note:
1. Population denotes the greater metro area population;
2. Source: All variables except Race and Greater Metro Area Characteristics (GMAC) extracted from the Lahman Baseball Database (Version 5.0, Release Date: Dec. 15, 2002). Race is derived form observed Topps Baseball Cards, years 92, 93, 94, 97, 99 (only years available). GMAC derived from the Statistical Abstract 1997-1999, the BEA, CA1-3, and from Statistical Canada

In Table 3, the domination of white pitchers is immediately apparent. White pitchers are on average older than both black and (especially) Hispanic pitchers. They also enjoy higher average earnings. In terms of career characteristics, white pitchers record significantly higher *Wins*, *Losses*, *Games Started*, *Complete Games*, *Shutouts*, *Saves*, *Homeruns*, *Walks*, *Strikeouts* and *Innings Pitched* than either blacks or Hispanic pitchers, with Hispanic pitchers recording generally lower figures than black pitchers.

3.2 *Empirical Methodology*

Wage discrimination occurs when individuals who are identical in terms of their productive characteristics are paid differently on account of their non-productive characteristics. Any empirical analysis of discrimination thus requires some control of productivity - it would not be surprising, and nor would it suggest discrimination, if more productive individuals were paid more than less productive individuals. In the traditional literature such control is usually straightforward since the individuals under scrutiny are performing the same job. In our model, however, it is problematic. Our concern is whether there is discrimination *across* job assignments, that is, where individuals with *different* non-productive characteristics are performing *different* jobs - do male airline pilots earn more than female flight attendants because of their occupation or because of their gender? This is a difficult issue to address empirically because we need to control for the productivity of both the pilot and the flight attendant or, more generally, we need to control for assignment-specific productivity. Clearly some measures of productivity will be common across job assignments - for example, education, job-tenure, and labour market experience. By definition, however, some measures of productivity will be unique to particular job assignments and it is controlling for these that is the real challenge.

One possible solution is to adopt a two-stage generated regressor approach.[9] Assume that wages reflect productivity as follows: To ascertain the level of discrimination across player positions, we need to control for position-specific productivity. In one sense this is straightforward because some measures of off-field productivity (MLB experience and tenure with current team, for example) are common across pitchers and hitters. On-field measures of productivity, however, vary across hitters and pitchers; e.g. runs for hitters and strike-outs for pitchers. Given our objective of ascertaining the extent of racial discrimination *across* job assignments, we need a standardized productivity measure. We thus adopt the following two-stage approach. We first assume that wages reflect productivity as follows:

$$\ln w^{ij} = \mathbf{X}_0^j \mathbf{B}_0^{ij} + \mathbf{X}_1 \mathbf{B}_1^{ij} \tag{5}$$

$\ln w^{ij}$ denotes the log wage of a member of group $i = 1, 2, ..., I$ employed in job assignment $j = 1, 2, ..., J$, $\mathbf{X}_0^j$ is a vector of 'assignment-specific' productivity measures, $\mathbf{X}_1$ is a vector of 'common' (i.e. cross assignment) productivity measures (e.g. education, tenure), and the **B**'s denote parameter vectors. Our aim is to derive an estimating equation of the form:

$$\ln w^{ij} = \mathbf{X}_0 \mathbf{B}_0^{ij} + \mathbf{X}_1 \mathbf{B}_1^{ij} \tag{6}$$

where $\mathbf{X}_0$ denotes some standardised (imputed) measure of assignment-specific productivity. To this end, we estimate the following 'first-stage' group-assignment regressions:

$$\ln w^{ij} = \mathbf{X}_0^j \mathbf{A}_0^{ij} \tag{7}$$

That is, we estimate separate wage regressions for each racial group employed within each job assignment, including as explanatory variables only each group's respective assignment-

[9] See Pagan (1984), Gauger (1989) and Gawande (1996) for discussions of the inference issues regarding estimated regressor models.

specific productivity measures. Thus, we estimate separate wage regressions for black, white and Hispanic hitters and pitchers on only their respective position-specific variables vis. Pitchers - *Starter*; *Wins*; *Losses*; *Games Started*; *Complete Games*; *Shutouts*; *Saves*; *Homeruns*; *Walks*; *Strikeouts*; *Innings Pitched*; *ERA*; and *Strikeout Rate*; Hitters – *At Bat*; *Stolen Bases*; *Bases on Balls*; *Total Bases*; *Slugging Average*; *Batting Average*; *Infielder*; *Outfielder*; *Catcher*; and *Designated Hitter*. We then use the predicted values from these regressions, $\hat{w} = \left(\hat{w}^{ij};\text{"}\ i, j\right)$, as a standardized measure of assignment-specific productivity in second-stage regressions of the form:

$$\ln w^{ij} = \hat{w}\mathbf{B}_0^{ij} + \mathbf{X}_1\mathbf{B}_1^{ij} \qquad (8)$$

3.3 *Cross-Assignment Regression Analysis*

Table 4 reports six second-stage regressions with white pitchers, black pitchers, Hispanic pitchers, white hitters, black hitters, and Hispanic hitters being defined as the default race-position category respectively.

The results in Table 4 show strong evidence of both cross- and within-assignment discrimination in MLB. Our estimated coefficients suggest that even after controlling for both on-and off-field productivity, white pitchers earn: (i) 16.4 per cent more than black pitchers; (ii) 17.0 per cent more than black hitters; (iii) 10.6 per cent more than black hitters; and (iv) 9.2 per cent (but only at the 90 per cent level of confidence) more than Hispanic hitters. Hispanic pitchers earn: (i) 17.0 per cent more than black pitchers; (ii) 17.6 per cent more than white hitters; and (iii) 11.2 per cent (but only at the 90 per cent level of confidence) more than black hitters.

We estimated a number of variants of the Table 4 regressions to test BO’s theoretical prior that discrimination increases with heightened customer prejudice, but can decline as labour markets become less competitive. Specifically, we re-estimated the Table 4 regressions for the ‘competitive’ and ‘non-competitive’ MLB markets separately, where the latter is defined as those players subject to the reserve clause or eligible for final offer arbitration. We also estimated separate Table 4 regressions for the ‘early’ (i.e. 1992 and 1993) and ‘latter’ (i.e. 1997 and 1998) periods of our data, both for the overall MLB market, and then for the competitive and non-competitive markets separately. Our objective here was to pick up the effects of the expansion in the size of the league, and the subsequent decline in monopsony power, during the 1990’s. Finally, we tested for customer discrimination generally, and the predictions from BO’s theoretical analysis particularly (i.e. that within a competitive labour market, an increase in customer prejudice will heighten the amount of wage discrimination across job assignments), by estimating separate Table 4 regressions for all players, ‘competitive’ players, and ‘non-competitive’ players, playing for teams located in greater metropolitan areas with above and below average non-white populations.

The results of these various regressions (over 80 in total) are available on request. For brevity we report the salient details only. We find discrimination to be generally more evident in the competitive MLB market than in the non-competitive MLB market, and also more evident in the ‘latter’ (i.e. post-expansion) period of our data than in the ‘early’ (i.e. ‘pre-expansion’) period. Breaking the analysis down further, discrimination appears to be more prevalent in the competitive MLB market in the latter period than it is in either the competitive market in the early period or the non-competitive market in the latter period, both of which exhibit more discrimination than the non-competitive market in the early period. In terms of customer discrimination, we find substantial evidence of discrimination in greater metropolitan areas with below average non-white populations, but less compelling evidence in those with above average non-white population. And finally, in terms of the former areas, discrimination appears to be more widespread in the competitive rather than in

Table 4: Discrimination Controlling for Position Specific Productivity

Dependent Variable: Log Annual Salary

	(1) All Default – White Pitcher		(2) All Default - Black Pitcher		(3) All Default – Hispanic Pitcher		(4) All Default – White Hitter		(5) All Default - Black Hitter		(6) All Default - Hispanic Hitter	
	Coef	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*
Imputed Productivity	0.863	34.05	0.863	34.05	0.863	34.05	0.863	34.05	0.863	34.05	34.05	34.05
Race Dummies												
White Pitcher	-	-	0.164	2.69	-0.005	-0.09	0.170	4.98	0.106	2.60	0.092	1.79
Black Pitcher	-0.164	-2.69	-	-	-0.170	-2.15	0.006	0.09	-0.058	-0.87	-0.072	-0.98
Hispanic Pitcher	0.005	0.09	0.170	2.15	-	-	0.176	2.81	0.112	1.70	0.097	1.35
White Hitter	-0.170	-4.98	-0.006	-0.09	-0.176	-2.81	-	-	-0.064	-1.52	-0.078	-1.46
Black Hitter	-0.106	-2.70	0.058	0.87	-0.112	-1.70	0.064	1.52	-	-	-0.014	-0.25
Hispanic Hitter	-0.092	-1.79	0.072	0.98	-0.097	-1.35	0.078	1.46	0.014	0.25	-	-
Professional Characteristics												
Age	-0.024	-3.27	-0.024	-3.27	-0.024	-3.27	-0.024	-3.27	-0.024	-3.27	-0.024	-3.27
MLB Experience	0.152	3.45	0.152	3.45	0.152	3.45	0.152	3.45	0.152	3.45	0.152	3.45
MLB Experience-Squared	-0.010	-4.79	-0.010	-4.79	-0.010	-4.79	-0.010	-4.79	-0.010	-4.79	-0.010	-4.79
Tenure	0.056	10.00	0.056	10.00	0.056	10.00	0.056	10.00	0.056	10.00	0.056	10.00
Free Agent	0.879	6.14	0.879	6.14	0.879	6.14	0.879	6.14	0.879	6.14	0.879	6.14
Final Offer Arbitration	0.471	5.94	0.471	5.94	0.471	5.94	0.471	5.94	0.471	5.94	0.471	5.94
American League	-0.006	-0.23	-0.006	-0.23	-0.006	-0.23	-0.006	-0.23	-0.006	-0.23	-0.006	-0.23
Canadian	-0.022	-0.21	-0.022	-0.21	-0.022	-0.21	-0.022	-0.21	-0.022	-0.21	-0.022	-0.21
Greater Metro Area Characteristics												
Per cent White	0.001	0.34	0.001	0.34	0.001	0.34	0.001	0.34	0.001	0.34	0.001	0.34
Per cent Black	0.005	1.24	0.005	1.24	0.005	1.24	0.005	1.24	0.005	1.24	0.005	1.24
Per cent Hispanic	0.005	3.39	0.005	3.39	0.005	3.39	0.005	3.39	0.005	3.39	0.005	3.39
Average Annual Income	0.000	1.45	0.000	1.45	0.000	1.45	0.000	1.45	0.000	1.45	0.000	1.45
Population	0.000	0.11	0.000	0.11	0.000	0.11	0.000	0.11	0.000	0.11	0.000	0.11
Year Dummies												
1993	0.051	1.31	0.051	1.31	0.051	1.31	0.051	1.31	0.051	1.31	0.051	1.31
1997	0.046	0.97	0.046	0.97	0.046	0.97	0.046	0.97	0.046	0.97	0.046	0.97
1998	0.130	2.44	0.130	2.44	0.130	2.44	0.130	2.44	0.130	2.44	0.130	2.44
Constant	0.994	1.63	0.830	1.37	0.100	1.65	0.824	1.34	0.888	1.44	0.902	1.48
R-Squared	0.7360		0.7360		0.7360		0.7360		0.7360		0.7360	
F-Statistic	$422.35_{22, 2274}$		$422.35_{22, 2274}$		$422.35_{22, 2274}$		$422.35_{22, 2274}$		$422.35_{22, 2274}$		$422.35_{22, 2274}$	
Root Mean Squared Error	0.61289		0.61289		0.61289		0.61289		0.61289		0.61289	
Observations	2297		2297		2297		2297		2297		2297	

the non-competitive MLB market. While the standard prediction regarding the relationship between discrimination and monopsony power is that of a positive relationship, our results appear to indicate a generally *negative* relationship. While this may seem counterintuitive, it is certainly consistent with BO's theoretical model, which is capable of predicting a negative relationship assuming certain parameter restrictions are in place.

In Table 5 we explore BO's theoretical prior that wage discrimination across player job assignments interacts with productivity differences between majority and minority workers. We test this prediction by creating a *Relative Productivity* variable that equals the difference between a player's individual productivity and the mean productivity of players in the other racial/position group multiplied by the player's individual productivity. Thus, in Column (1) of Table 5, where we focus on white pitchers relative to black hitters, our *Relative Productivity* (*White Pitcher:Black Hitter*) variable is defined as: *Individual White Pitcher Productivity* x (*Individual White Pitcher Productivity - Mean Black Hitter Productivity*), where productivity is estimated according to the two-stage process outlined in equations (5)-(8).

There is some tentative evidence from Tables 5 that relative productivity does affect *ceteris paribus* race-position salary differentials. Our empirical results suggest that whilst the white hitter / black pitcher, black hitter / white pitcher and black hitter / Hispanic pitcher wage differentials are unaffected by relative productivity differences, the differential of white and Hispanic hitters over black pitchers increases with the relative productivity of the former. Moreover, the differential of white pitchers over Hispanic Hitters declines with increases in the relative productivity of the latter.

Table 5: Discrimination Controlling for Position Specific Productivity and Relative Productivity (Hitters – Pitchers)
Dependent Variable: Log Annual Salary

	(1) White Hitters / Black Pitchers		(2) White Hitters / Hispanic Pitchers		(3) Black Hitters / White Pitchers		(4) Black Hitters / Hispanic Pitchers		(5) Hispanic Hitters / White Pitchers		(6) Hispanic Hitters / Black Pitchers	
	Coef	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*	*Coef*	*T-Stat*
Imputed Productivity	0.635	10.38	0.950	10.27	0.862	23.55	1.012	10.48	0.838	22.05	0.730	10.17
Race Dummies												
White Hitters	-0.030	-0.47	-0.189	-2.75	-	-	-	-	-	-	-	-
Black Hitters	-	-	-	-	-0.109	-2.57	-0.168	-2.24	-	-	-	-
Hispanic Hitters	-	-	-	-	-	-	-	-	-0.137	-2.42	0.052	0.70
Relative Productivity												
White Hitter: Black Pitcher	0.018	4.12	-	-	-	-	-	-	-	-	-	-
White Hitter: Hispanic Pitcher	-	-	-0.005	-0.84	-	-	-	-	-	-	-	-
Black Hitter: White Pitcher	-	-	-	-	-0.000	-0.11	-	-	-	-	-	-
Black Hitter: Hispanic Pitcher	-	-	-	-	-	-	-0.004	-0.65	-	-	-	-
Hispanic Hitter: White Pitcher	-	-	-	-	-	-	-	-	0.007	1.90		
Hispanic Hitter: Black Pitcher	-	-	-	-	-	-	-	-	-	-	0.018	3.65
Constant	3.386	2.60	-0.211	-0.13	1.346	1.67	-1.800	-1.05	1.511	1.82	-0.096	-0.06
R-Squared	0.7200		0.7727		0.7481		0.7347		0.7620		0.7840	
F-Statistic	$132.29_{19, 656}$		$132.82_{19, 664}$		$321.88_{19, 1289}$		$104.98_{19, 482}$		$309.12_{19, 1099}$		$101.52_{19, 284}$	
Root Mean Squared Error	0.61053		0.61184		0.60507		0.61840		0.59513		0.59260	
Observations	676		684		1309		502		1119		304	

Notes:
1. Other explanatory regressors were those set out in Table 3;
2. 'Relative Productivity' is defined as, e.g., 'White Hitter: Black Pitcher = Individual White Hitter Productivity x (Individual White Hitter Productivity - Mean Black Pitcher Productivity).

3.4. Decomposition Analysis

In this section, we attempt to identify cross-assignment discrimination using another empirical approach. The fact that players of a particular race in a particular position enjoy a wage differential over players of another race in another position could be a reflection of the former group's greater endowment of 'earning characteristics'. White pitchers may, for example, be more productive or have more experience on average than non-white (i.e. black or Hispanic) hitters. Alternatively, white pitchers may be better rewarded for the characteristics they do possess, suggesting some form of positive (negative) discrimination from employers towards white pitchers (non-white hitters). To address this issue we perform a Blinder-Oaxaca *decomposition* to separate the earnings differential into an 'endowment component', to account for differences in endowments between individuals, and a 'price component', which is usually associated with discrimination.[10]

Recalling equation (7), we write the earnings function of players of race *j* in position *i* as:

$$\ln w^{ij} = \mathbf{X}^{ij}\mathbf{B}^{ij} + \varepsilon^{ij} \tag{9}$$

where $i = (W, NW)$ and $j = (H, P)$ denote white and non-white and pitchers and hitters respectively, and where $NW = (B, H)$ denotes black and Hispanic respectively. $\mathbf{X}^{ij} = \left(\mathbf{X}_0^{ij}, \mathbf{X}_1^{ij}\right)$ denotes our vectors of position-specific and common productivity characteristics, $\mathbf{B}^{ij} = \left(\mathbf{B}_0^{ij}, \mathbf{B}_1^{ij}\right)$ the corresponding coefficient vectors to be estimated, and e^{ij} some well-behaved error term. Thus, the earnings functions of white pitchers, non-white pitchers, white hitters and non-white hitters may be denoted:

[10] This method of decomposition, initially proposed by Oaxaca (1973) and Blinder (1973), and later generalized by Oaxaca and Ransom (1994), has been applied extensively to discrimination on the basis of gender, race, caste and religion.

$$\ln w^{WP} = \mathbf{X}^{WP}\mathbf{B}^{WP} + \varepsilon^{WP} \tag{10}$$

$$\ln w^{NWP} = \mathbf{X}^{NWP}\mathbf{B}^{NWP} + \varepsilon^{NWP} \tag{11}$$

$$\ln w^{WH} = \mathbf{X}^{WH}\mathbf{B}^{WH} + \varepsilon^{WH} \tag{12}$$

$$\ln w^{NWH} = \mathbf{X}^{NWH}\mathbf{B}^{NWH} + \varepsilon^{NWH} \tag{13}$$

The Blinder-Oaxaca decomposition divides wage differentials into a part that is ‘explained’ by group differences in productivity and a residual part that cannot be accounted for by such differences in wage determinants. This latter ‘unexplained’ component is often used as a measure for discrimination. For example, the predicted average white pitcher/non-white hitter (WP-NWH) differential may be represented as:

$$\begin{gathered}\Delta \ln w^{WP-NWH} = \ln w^{WP} - \ln w^{NWH} = \bar{\mathbf{X}}^{WP}\hat{\mathbf{B}}^{WP} - \bar{\mathbf{X}}^{NWH}\hat{\mathbf{B}}^{NWH} \\ \Rightarrow \\ \Delta \ln w^{WP-NWH} = \hat{\mathbf{B}}^{NWH}\left(\bar{\mathbf{X}}^{WP} - \bar{\mathbf{X}}^{NWH}\right) + \bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right)\end{gathered} \tag{14}$$

The first term, $\hat{\mathbf{B}}^{NWH}\left(\bar{\mathbf{X}}^{WP} - \bar{\mathbf{X}}^{NWH}\right)$, represents differences in endowments between members of the two groups whilst the second term, $\bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right)$, represents differences in rewards. Note that if the overall differential is negative (i.e. $\Delta \ln w^{WP-NWH} < 0$) but the second term is positive [i.e. $\bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right) > 0$], then it would suggest that non-white hitters are discriminated against despite earning, on average, more than white hitters - i.e. non-white hitters would do even better with the earnings generating function of white pitchers than with their own.

Specification (14) presumes that the non-white hitter wage structure prevails in the absence of discrimination. But this is a matter of debate. Assuming away any feelings of malevolence or benevolence from one group towards the other, then it is equally valid to presume that the white pitcher wage structure prevails, thereby requiring (14) to be re-specified as:

$$\Delta \ln w^{WP-NWH} = \hat{\mathbf{B}}^{WP}\left(\bar{\mathbf{X}}^{WP} - \bar{\mathbf{X}}^{NWH}\right) + \bar{\mathbf{X}}^{NWH}\left(\hat{\mathbf{B}}^{WP} - \hat{\mathbf{B}}^{NWH}\right) \tag{15}$$

The first and second terms on the right hand side of (14) still represent differences in endowments and rewards respectively, but they will generally differ from those derived from equation (13).[11] Many authors concede this ambiguity by simply reporting both decompositions. Some, however, have attempted to confront the issue head-on by hypothesizing the non-discriminatory parameter vector, $\bar{\mathbf{B}}$, directly.[12] Reimers (1983), for example, proposes using the average coefficients over both groups as an estimate of $\bar{\mathbf{B}}$. Neumark (1988) advocates using the coefficients from a pooled regression over both groups as an estimate of $\bar{\mathrm{B}}$. In what follows, we follow the 'hybrid' decomposition technique popularized by Cotton (1988) in which the prevailing non-discriminatory wage structure is assumed to be a weighted average of the wage structures of the two groups under consideration:

$$\Delta \ln w^{WP-NWH} = \bar{\mathbf{X}}^{WP}\left(\hat{\mathbf{B}}^{WP} - \bar{\mathbf{B}}\right) + \bar{\mathbf{X}}^{NWH}\left(\bar{\mathbf{B}} - \hat{\mathbf{B}}^{NWH}\right) + \bar{\mathbf{B}}\left(\bar{\mathbf{X}}^{WP} - \bar{\mathbf{X}}^{NWH}\right) \tag{16}$$

where $\bar{\mathbf{B}} = \mathbf{\Omega}\hat{\mathbf{B}}^{WP} + \left(1 - \mathbf{\Omega}\right)\hat{\mathbf{B}}^{NWH}$ represents the estimated non-discriminatory parameter vector, with $\mathbf{\Omega}$ denoting the proportion of the sample comprised by white pitchers. The first right-hand

[11] The point that an undervaluation of one group implies an overvaluation of the other is neatly summarized by Cotton (1988, p. 238): '… not only is the group discriminated against undervalued, but the preferred group is overvalued, and the undervaluation of the one subsidizes the overvaluation of the other.'

[12] Oaxaca and Ransom (1994) provide an integrative treatment of the various methods.

term in the decomposition is the overpayment enjoyed by white pitchers, the second term is the underpayment suffered by non-white hitters, and the third term is the portion of the wage differential that is explained by differences in endowments. We perform the above three decompositions for the white pitcher/non-white hitter and white hitter/non-white pitcher differentials, and our results, based on the regressions set out in Table 4, are collected in Tables 6a-6d.

Considering Table 6a, our regression model implies a positive salary premium for black hitters over white pitchers *ceteris paribus*. The first decomposition, which follows specification (14) in presuming the black hitter wage structure would prevail in the absence of any discrimination, suggests that this premium would be even greater in the absence of discrimination, with discrimination *against* black hitters alleviating the potential differential by some 33 percent. The second decomposition, which follows specification (15) in presuming that the white pitcher wage structure would prevail in the absence of discrimination, suggests that discrimination against black hitters alleviates the overall potential differential by a somewhat less, but still considerable, 22 percent. The hybrid decomposition, derived from specification (16), echoes the finding that discrimination assuages the potential black hitter wage premium with white pitcher overpayment and black hitter underpayment reducing the potential premium by approximately 9 per cent and 15 per cent respectively.

Table 6b focuses on the white pitcher / Hispanic hitter differential. Our results here imply a positive salary premium for Hispanic hitters over white pitchers *ceteris paribus*. The decomposition of this differential suggests even larger discrimination than that evident in the white pitcher / black hitter differential. Decomposition based on the white pitcher wage structure suggests that discrimination against Hispanic hitters reduces the potential Hispanic hitter

Table 6a: Oaxaca-Cotton Decompositions: White Pitcher / Black Hitter

$$\Delta \ln w^{WP-BH} = \ln w^{WP} - \ln w^{BH}$$

		Coef.	*%*
Black Hitter Wage Structure			
Endowment Effect:	$\hat{B}^{BH}\left(\bar{X}^{WP}-\bar{X}^{BH}\right)$	-0.649	*133.29*
Price Effect:	$\bar{X}^{WP}\left(\hat{B}^{WP}-\hat{B}^{BH}\right)$	0.162	*-33.29*
Total Differential:	$\hat{B}^{BH}\left(\bar{X}^{WP}-\bar{X}^{BH}\right)+\bar{X}^{WP}\left(\hat{B}^{WP}-\hat{B}^{BH}\right)$	-0.487	*100.00*
White Pitcher Wage Structure			
Endowment Effect:	$\hat{B}^{WP}\left(\bar{X}^{WP}-\bar{X}^{BH}\right)$	-0.591	*121.49*
Price Effect:	$\bar{X}^{BH}\left(\hat{B}^{WP}-\hat{B}^{BH}\right)$	0.104	*-21.49*
Total Differential:	$\hat{B}^{WP}\left(\bar{X}^{WP}-\bar{X}^{BH}\right)+\bar{X}^{BH}\left(\hat{B}^{WP}-\hat{B}^{BH}\right)$	-0.487	*100.00*
Hybrid Wage Structure			
White Pitcher Overpayment:	$\bar{X}^{WP}\left(\hat{B}^{WP}-\bar{B}\right)$	0.045	*-9.33*
Black Hitter Underpayment:	$\bar{X}^{BH}\left(\bar{B}-\hat{B}^{BH}\right)$	0.075	*-15.47*
Endowment Effect:	$\bar{B}\left(\bar{X}^{WP}-\bar{X}^{BH}\right)$	-0.607	*124.80*
Total Differential:	$\bar{X}^{WP}\left(\hat{B}^{WP}-\bar{B}\right)+\bar{X}^{BH}\left(\bar{B}-\hat{B}^{BH}\right)+\bar{B}\left(\bar{X}^{WP}-\bar{X}^{BH}\right)$	-0.487	*100.00*

Table 6b: Oaxaca-Cotton Decompositions: White Pitcher / Hispanic Hitter

$$\Delta \ln w^{WP-HH} = \ln w^{WP} - \ln w^{HH}$$

		Coef.	*%*
Hispanic Hitter Wage Structure			
Endowment Effect:	$\hat{B}^{HH}\left(\bar{X}^{WP}-\bar{X}^{HH}\right)$	-0.604	*145.33*
Price Effect:	$\bar{X}^{WP}\left(\hat{B}^{WP}-\hat{B}^{HH}\right)$	0.189	*-45.33*
Total Differential:	$\hat{B}^{HH}\left(\bar{X}^{WP}-\bar{X}^{HH}\right)+\bar{X}^{WP}\left(\hat{B}^{WP}-\hat{B}^{HH}\right)$	-0.416	*100.00*
White Pitcher Wage Structure			
Endowment Effect:	$\hat{B}^{WP}\left(\bar{X}^{WP}-\bar{X}^{HH}\right)$	-0.512	*123.12*
Price Effect:	$\bar{X}^{HH}\left(\hat{B}^{WP}-\hat{B}^{HH}\right)$	0.096	*-23.12*
Total Differential:	$\hat{B}^{WP}\left(\bar{X}^{WP}-\bar{X}^{HH}\right)+\bar{X}^{HH}\left(\hat{B}^{WP}-\hat{B}^{HH}\right)$	-0.416	*100.00*
Hybrid Wage Structure			
White Pitcher Overpayment:	$\bar{X}^{WP}\left(\hat{B}^{WP}-\bar{B}\right)$	0.030	*-7.17*
Hispanic Hitter Underpayment:	$\bar{X}^{HH}\left(\bar{B}-\hat{B}^{HH}\right)$	0.081	*-19.46*
Endowment Effect:	$\bar{B}\left(\bar{X}^{WP}-\bar{X}^{HH}\right)$	-0.527	*126.63*
Total Differential:	$\bar{X}^{WP}\left(\hat{B}^{WP}-\bar{B}\right)+\bar{X}^{HH}\left(\bar{B}-\hat{B}^{HH}\right)+\bar{B}\left(\bar{X}^{WP}-\bar{X}^{HH}\right)$	-0.416	*100.00*

Table 6c: Oaxaca-Cotton Decompositions: White Hitter / Black Pitcher

$\Delta \ln w^{WH-BP} = \ln w^{WH} - \ln w^{BP}$

		Coef.	*%*
Black Pitcher Wage Structure			
Endowment Effect:	$\hat{B}^{BP}\left(\bar{X}^{WH} - \bar{X}^{BP}\right)$	0.660	*105.27*
Price Effect:	$\bar{X}^{WH}\left(\hat{B}^{WH} - \hat{B}^{BP}\right)$	-0.033	*-5.27*
Total Differential:	$\hat{B}^{BP}\left(\bar{X}^{WH} - \bar{X}^{BP}\right) + \bar{X}^{WH}\left(\hat{B}^{WH} - \hat{B}^{BP}\right)$	0.627	*100.00*
White Hitter Wage Structure			
Endowment Effect:	$\hat{B}^{WH}\left(\bar{X}^{WH} - \bar{X}^{BP}\right)$	0.639	*101.89*
Price Effect:	$\bar{X}^{BP}\left(\hat{B}^{WH} - \hat{B}^{BP}\right)$	-0.012	*-1.89*
Total Differential:	$\hat{B}^{WH}\left(\bar{X}^{WH} - \bar{X}^{BP}\right) + \bar{X}^{BP}\left(\hat{B}^{WH} - \hat{B}^{BP}\right)$	0.627	*100.00*
Hybrid Wage Structure			
White Hitter Overpayment:	$\bar{X}^{WH}\left(\hat{B}^{WH} - \bar{B}\right)$	-0.006	*-0.99*
Black Pitcher Underpayment:	$\bar{X}^{BP}\left(\bar{B} - \hat{B}^{BP}\right)$	-0.010	*-1.53*
Endowment Effect:	$\bar{B}\left(\bar{X}^{WH} - \bar{X}^{BP}\right)$	0.643	*102.52*
Total Differential:	$\bar{X}^{WH}\left(\hat{B}^{WH} - \bar{B}\right) + \bar{X}^{BP}\left(\bar{B} - \hat{B}^{BP}\right) + \bar{B}\left(\bar{X}^{WH} - \bar{X}^{BP}\right)$	0.627	*100.00*

Table 6d: Oaxaca-Cotton Decompositions: White Hitter / Hispanic Pitcher

$\Delta \ln w^{WH-HP} = \ln w^{WH} - \ln w^{HP}$

		Coef.	*%*
Hispanic Pitcher Wage Structure			
Endowment Effect:	$\hat{B}^{HP}\left(\bar{X}^{WH} - \bar{X}^{HP}\right)$	0.859	*145.76*
Price Effect:	$\bar{X}^{WH}\left(\hat{B}^{WH} - \hat{B}^{HP}\right)$	-0.270	*-45.76*
Total Differential:	$\hat{B}^{HP}\left(\bar{X}^{WH} - \bar{X}^{HP}\right) + \bar{X}^{WH}\left(\hat{B}^{WH} - \hat{B}^{HP}\right)$	0.589	*100.00*
White Hitter Wage Structure			
Endowment Effect:	$\hat{B}^{WH}\left(\bar{X}^{WH} - \bar{X}^{HP}\right)$	0.765	*129.86*
Price Effect:	$\bar{X}^{HP}\left(\hat{B}^{WH} - \hat{B}^{HP}\right)$	-0.176	*-29.86*
Total Differential:	$\hat{B}^{WH}\left(\bar{X}^{WH} - \bar{X}^{HP}\right) + \bar{X}^{HP}\left(\hat{B}^{WH} - \hat{B}^{HP}\right)$	0.589	*100.00*
Hybrid Wage Structure			
White Hitter Overpayment:	$\bar{X}^{WH}\left(\hat{B}^{WH} - \bar{B}\right)$	-0.053	*-9.03*
Hispanic Pitcher Underpayment:	$\bar{X}^{HP}\left(\bar{B} - \hat{B}^{HP}\right)$	-0.141	*-23.96*
Endowment Effect:	$\bar{B}\left(\bar{X}^{WH} - \bar{X}^{HP}\right)$	0.784	*132.99*
Total Differential:	$\bar{X}^{WH}\left(\hat{B}^{WH} - \bar{B}\right) + \bar{X}^{HP}\left(\bar{B} - \hat{B}^{HP}\right) + \bar{B}\left(\bar{X}^{WH} - \bar{X}^{HP}\right)$	0.589	*100.00*

premium by over 45 percent, whilst decomposition based on the Hispanic hitter wage structure puts the figure at 23 percent. The hybrid decomposition suggests that white pitcher overpayment and Hispanic hitter underpayment offset the potential Hispanic hitter wage premium by approximately 7 per cent and 19 per cent respectively.

Tables 6c and 6d focus on the white hitter / black pitcher and white hitter / Hispanic pitcher decomposition. Both decompositions imply a positive salary premium for white hitters. Table 6c suggests that discrimination plays a relative minor role in the white hitter / black pitcher differential, discrimination *against* white hitters reducing the potential white hitter premium by just over 5 per cent according to the black pitcher wage structure, and just under 2 per cent according to the white hitter wage structure. The hybrid decomposition implies white hitter overpayment and black pitcher underpayment reduce the differential by 1 per cent and 1.5 per cent respectively.

It would appear that discrimination plays a much more significant role in the white hitter / Hispanic pitcher differential. According to Table 6d, discrimination *against* white hitters reduces the potential differential by 46 per cent according to the Hispanic pitcher wage structure and by 30 per cent according to the white hitter wage structure. The hybrid decomposition suggests that white hitter overpayment and Hispanic pitcher underpayment reduces the potential white hitter premium by 9 per cent and 24 per cent respectively.

6. Final Comments

In this study, we address a widely neglected problem in the literature on taste discrimination in pay: Ascertaining the extent to which racial or gender differences in pay across job assignments are attributable to prejudice. Nearly all wage discrimination studies have focused on discrimination within the same job assignment, thus treating whites and non-whites (or males and females) as perfect substitutes. In a recent contribution, BO extend the theory to the case of discrimination across job assignments where assignments are viewed as distinct inputs. BO's theoretical findings underscore the importance of carefully considering the production function when there are productivity differences between majority and minority workers. An important finding from our theoretical analysis is that the magnitude of white/non-white productivity differences influences the amount of discrimination. Furthermore, when whites and non-whites are interrelated in production, race and productivity will interact. This is an important implication, for it means that *whenever* white and non-white workers have productivity differences, the researcher should include productivity x race interactions in any empirical specification.

We test BO's model using data from Major League Baseball, an industry characterized by complementary job assignments, a history of racial integration and discrimination, and a dual labour market structure. We find convincing evidence of racial differences in pay across player job assignments, even after controlling for a wide array of demographic variables and position-specific productivity. Moreover, we find strong evidence of BO's theoretical prior that racial pay differentials across assignments are affected by changes in relative productivities.

References

Arrow, K. J. (1973). 'The Theory of Discrimination.' In O.A. Ashenfelter and A. Rees (Eds.), *Discrimination in Labour Markets*. Princeton NJ: Princeton University Press, pp. 3-33.

Becker, G. S. (1971). *The Economics of Discrimination, Second Edition.* Chicago: University of Chicago Press.

Blinder, A.S. (1973). 'Wage Discrimination: Reduced Form and Structural Estimates.' *Journal of Human Resources*, 2, pp. 8-22.

Bodvarsson, Ö. B. and J. G. Sessions. (2011). 'The Measurement of Pay Discrimination Between Job Assignments.' *Labour Economics*. 18 pp. 645-657.

Bodvarsson, Ö. B. and M. D. Partridge. (2001). 'A Supply and Demand Model of Employer, Customer and Co-worker Discrimination.' *Labour Economics*, 8, June, pp. 389-416.

Borjas, G. J. (1983). 'The Substitutability of Black, Hispanic, and White Labour.' *Economic Inquiry*, 21, January, pp. 93-106.

Borjas, G. J. (1987). 'Immigrants, Minorities, and Labour Market Competition.' *Industrial and Labour Relations Review*, 40 (3), April, pp. 382-92.

Borjas, G. J. (2008). *Labour Economics*, 4th edition. Dubuque, IA: McGraw-Hill Irwin.

Cotton, J. (1988). 'On the Decomposition of Wage Differentials.' *Review of Economics and Statistics*, 70, pp. 236-243.

Gauger, J. (1989). The Generated Regressor Correction: Impacts upon Inferences in Hypothesis Testing.' *Journal of Macroeconomics*, 11(3), pp.383-395.
Gawande, K. (1997). Generated Regressors in Linear and Nonlinear Models.' *Economics Letters*, 54, pp119-126.
Grant, J.H. and D.S. Hamermesh (1981). 'Labour Market Competition Among Youths, White Women and Others.' *The Review of Economics and Statistics*, 63(3), August, pp. 354-60.
Hashimoto, M. and L. Kochin (1980). 'A bias in the statistical effects of discrimination.' *Economic Inquiry*, 8, pp. 478-86.
Kahanec, M. (2006). 'The Substitutability of Labour of Selected Ethnic Groups in the U.S. Labour Market.' *IZA Discussion Paper No. 1945*, January (www.iza.org).
Kahn, L. M. (1991). 'Customer Discrimination and Affirmative Action.' *Economic Inquiry*, 29, July, pp. 555-571.
Kahn, L. M. (2000). 'A Level Playing Field? Sports and Discrimination.' In William S. Kern (ed.), *The Economics of Sports*, W.E. Upjohn Institute, Kalamazoo, MI.
Neumark, D. (1988). 'Employers' Discriminatory Behavior and the Estimation of Wage Discrimination.' *The Journal of Human Resources*, 23, pp.279–295.
Oaxaca, R. L. (1973). 'Male-Female Wage Differentials in Urban Labour Markets.' *International Economic Review*, 14, pp. 337-356.
Oaxaca, R. L. and M. R. Ransom. (1994) 'On Discrimination and the Decomposition of Wage Differentials.' *Journal of Econometrics*, 61(1), pp. 5-21.
Pagan, A. (1984). 'Econometric Issues in the Analysis of Regressions with Generated Regressors.' *International Economic Review*, 25(1), pp. 221-247
Reimers, C. W. (1983). 'Labour Market Discrimination against Hispanic and Black Men.' *The Review of Economics and Statistics*, 65, pp. 570–579.
Welch, F. (1967). 'Labour-Market Discrimination: An Interpretation of Income Differences in the Rural South.' *Journal of Political Economy*, 75(3), June, pp. 225-40.
Woolway, M.D. (1997). 'Using an Empirically Estimated Production Function for Major League Baseball to Examine Worker Disincentives Associated with Multi-Year Contracts.' *The American Economist*, 41(2), 77-84.
Zech, C.E. (1981). 'An Empirical Estimation of a Production Function: The Case of Major League Baseball.' *The American Economist*, 25(2), 19-23.

Appendix: Baseball Terminology

1. A player has an *at bat* every time he comes to bat, except in certain circumstances, e.g. if he is awarded first base due to interference or obstruction or the inning ends while he is still at bat.

2. A hitter is assigned a *stolen base* (also called a *steal*) when he reaches an extra base on a hit from another player. For example, suppose that hitter A is at first base when hitter B hits the ball. Hitter B reaches first base (he would be assigned a *single*), but hitter A reaches third base. Hitter A would be assigned a *stolen base* because he reached an extra base.

3. A *base on balls* (also called a *walk*) is assigned when the batter receives four pitches each of which the umpire determines is a *ball*. A *ball* is any pitch at which the batter does not swing and is out of the *strike zone* (which means it would not qualify to be a *strike*). When the hitter is assigned a *base on balls*, he is entitled to walk to first base.

4. *Total bases* are the number of bases a player has gained through hitting. It is the sum of his *hits* weighted by 1 for a *single*, 2 for a *double* (if he gets to second base as a result of his hit), 3 for a *triple* (if he gets to third base) and 4 for a *home run*.

5. A hitter's *batting average* is the ratio of *hits* to *at bats*; this measures the hitter's success rate. *Slugging percentage*, a related measure, reflects hitting power, which is *total bases* divided by *at bats*.

6. An *infielder* is a defensive player who plays on the *infield*, the dirt portion of a baseball diamond between first and third bases. The specific infielder positions are *first baseman*, *second baseman*, *shortstop* (which is between second and third bases) and *third baseman*. In contrast, an *outfielder* plays farthest from the batter and his primary role is to catch long fly balls. Outfielder positions include *left fielder*, *center fielder* and *right fielder*. The *catcher* crouches behind home plate and receives the ball from the *pitcher*. Because the *catcher* can see the whole field, he is best positioned to lead and direct his fellow players in play. He typically calls the pitches by means of hand signals, hence requires awareness of both the pitcher's mechanics and the strengths and weaknesses of the batter.

7. A *pitcher* is assigned a *win* or a *loss* depending on whether he was the *pitcher of record* when the decisive run was scored. One is the *pitcher of record* if one is the *pitcher* at the point when the player who scores the decisive run is allowed to reach a base.

8. *Games started* is the number of times the *pitcher* was given the ball to start a game, whereas *games finished* is the number of times the pitcher was throwing on the mound during the final *out* (which is any failed attempt by a hitter to advance to a base).

9. A *shutout* is a game in which one team does not score any runs. A pitcher earns a *save* if he is able to hold a lead for his team at the end of the game.

10. *Pitchers* who earn saves, called *relievers*, tend not to gain *wins*, so it is customary to treat *saves* and *wins* equally, especially when studying *pitcher* salaries.

11. Number of *home runs*, which is assumed to be negatively related to salary, is the number of pitches that were hit by batters which were scored as a *home run*.

12. A *pitcher* is assigned a *walk*, which is assumed to be negatively related to salary, if he allows a batter to reach base after pitching him four balls. He is assigned a *strikeout* if he pitches three *strikes* (pitched balls counted against the batter, typically swung at and missed or fouled off) in a row.

13. An *inning* is one of nine periods in a MLB game in which each team has a turn at bat; *innings pitched* is the number of such periods when the pitcher was working.

14. *Earned run average* is negatively correlated with the pitcher's ability to prevent the opposing team from scoring. It equals the number of times the pitcher allows a batter to score a *run* (where the batter scores a point by advancing around the bases and reaching home plate safely) x 9, divided by the number of *innings pitched*.

15. The *strikeout rate* is the percentage of times the pitcher has succeeded in striking a batter out.