

The examination of pay discrimination based on race in the MLB

MATTHEW NAY



Intro and Motivation

The current makeup of players in the MLB in 2022:

- White - 62.1%
- Hispanic or Latino - 28.5%
- Black or African American - 7.2%
- Asian - 1.9%
- Hawaiian or Pacific Islander - 0.3%
- American Indian or Alaska Native - 0.1%

Been studied before and nothing has been found:

- Medoff (1975)
- Christiano (1986, 1988)
- Kahn (2000)
- King and Palmer (2006)

Data Set Details

At its largest:

- Hitters: 587 observations, 26 variables
- Pitchers: 286 observations, 31 variables

Data decisions:

- Need to be eligible to qualify for season awards
- Need to have signed a contract, can't be making league minimum

Table 1: Variable Definitions

Variable	Definition
Team	Team for that year
ln(Salary)	The Natural Log of each observations salary for that year
WAR	Wins Above Replacement
Age	Player's age that season
Service Time	Years player has played in the MLB
NonWhite	1 if player is not white, 0 otherwise
International	1 if player is not from USA, 0 otherwise
Year	year
Traditional Stats	
Hits	Number of hits collected by player
Home Runs	Number of home runs collected by player
Strikeouts	Number of strikeouts collected by player
Walks	Number of walks collected by player
AVG	Batting average by player
SLG	Slugging percentage
OBP	On base percentage
OPS	On base percentage plus slugging
Statcast Stats	
xBA	Expected batting average
xSLG	Expected slugging percentage
xwOBA	Expected weighted on base average
xOBP	Expected on base percentage
xISO	Expected isolated power
Avg EV	Average exit velo of baseball
Avg LA	Average launch angle of baseball
Sweet Spot %	Percent of at bats that are a batted-ball event with a launch angle between eight and 32 degrees
Barrel %	Percent of at bats that are a batted ball with the perfect combination of exit velocity and launch angle

First Model

$$\ln(\text{Salary}) = \beta_0 + \beta_1 * \text{WAR} + \beta_2 * \text{age} + \beta_3 * \text{NonWhite} + \beta_4 * \text{year}$$

	(1)
	ln_salary
war	0.0711*** (5.19)
age	0.141*** (15.22)
nonwhite	-0.00272 (-0.05)
year	0.0220 (1.84)
servicetime	
Constant	-32.99 (-1.36)
Traditional Stats	No
Sabermetric Stats	No
Team Fixed Effects	No
<i>N</i>	587
<i>R</i> ²	0.2872

Seventh Model

$$\ln(\text{Salary}) = \beta_0 + \beta_1 * \text{WAR} + \beta_2 * \text{NonWhite} + \beta_3 * \text{ServiceTime} + \sum_i \beta_{i+3} * \text{Trad}_i + \sum_j \beta_{I+j+3} * \text{Sabr}_i + \text{Team F.E.}$$

	(7) ln_salary
war	0.0358 (1.72)
age	
nonwhite	-0.169** (-3.11)
year	
servicetime	0.174*** (19.28)
Constant	10.18*** (5.42)
Traditional Stats	Yes
Sabermetric Stats	Yes
Team Fixed Effects	Yes
<i>N</i>	535
<i>R</i> ²	0.6103

Effect of NonWhite on ln(Salary) - Hitters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary
war	0.0711*** (5.19)	0.0688*** (5.96)	0.0369* (2.41)	0.0366* (2.39)	0.0389 (1.92)	0.0438** (2.85)	0.0358 (1.72)
age	0.141*** (15.22)						
nonwhite	-0.00272 (-0.05)	-0.188*** (-3.81)	-0.176*** (-3.43)	-0.176*** (-3.43)	-0.150** (-2.77)	-0.158** (-2.95)	-0.169** (-3.11)
year	0.0220 (1.84)	0.0322** (3.10)		0.00557 (0.45)			
servicetime		0.180*** (22.56)	0.175*** (21.40)	0.175*** (21.38)	0.173*** (19.24)	0.176*** (20.46)	0.174*** (19.28)
Constant	-32.99 (-1.36)	-50.43* (-2.41)	10.36*** (6.21)	-0.836 (-0.03)	14.13*** (29.53)	9.500*** (5.32)	10.18*** (5.42)
Traditional Stats	No	No	No	No	Yes	No	Yes
Sabermetric Stats	No	No	Yes	Yes	No	Yes	Yes
Team Fixed Effects	No	No	No	No	Yes	Yes	Yes
<i>N</i>	587	587	587	587	535	535	535
<i>R</i> ²	0.2872	0.4685	0.5068	0.5069	0.5794	0.6005	0.6103

t statistics in parentheses

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

Effect of International on ln(Salary) - Hitters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary
war	0.0689*** (5.02)	0.0669*** (5.78)	0.0333* (2.18)	0.0329* (2.15)	0.0322 (1.58)	0.0401** (2.60)	0.0285 (1.37)
age	0.141*** (15.20)						
international	-0.0695 (-1.15)	-0.211*** (-4.03)	-0.200*** (-3.79)	-0.202*** (-3.81)	-0.154** (-2.61)	-0.172** (-2.97)	-0.182** (-3.09)
year	0.0220 (1.84)	0.0321** (3.11)		0.00750 (0.60)			
servicetime		0.179*** (22.56)	0.174*** (21.47)	0.174*** (21.45)	0.172*** (19.14)	0.175*** (20.47)	0.171*** (19.19)
Constant	-32.77 (-1.36)	-50.20* (-2.40)	10.32*** (6.22)	-4.759 (-0.19)	14.15*** (29.37)	9.313*** (5.25)	9.957*** (5.33)
Traditional Stats	No	No	No	No	Yes	No	Yes
Sabermetric Stats	No	No	Yes	Yes	No	Yes	Yes
Team Fixed Effects	No	No	No	No	Yes	Yes	Yes
<i>N</i>	587	587	587	587	535	535	535
<i>R</i> ²	0.2889	0.4700	0.5090	0.5093	0.5787	0.6006	0.6102

t statistics in parentheses

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

Effect of NonWhite on ln(Salary) - Pitchers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary	ln_salary
war	0.0682** (3.28)	0.0632** (3.24)	0.0283 (0.79)	0.0283 (0.78)	0.00382 (0.07)	0.0218 (0.51)	0.0153 (0.26)
age	0.108*** (8.72)						
nonwhite	-0.111 (-1.22)	-0.125 (-1.47)	-0.0457 (-0.53)	-0.0453 (-0.52)	-0.0494 (-0.56)	-0.0216 (-0.24)	-0.00828 (-0.09)
year	-0.00512 (-0.28)	0.00281 (0.16)		-0.00175 (-0.08)			
servicetime		0.132*** (11.09)	0.134*** (11.21)	0.134*** (11.17)	0.122*** (10.00)	0.124*** (10.23)	0.124*** (9.74)
Constant	22.84 (0.62)	9.108 (0.26)	19.61*** (5.35)	23.16 (0.49)	14.26*** (11.55)	19.49*** (5.05)	18.92*** (4.51)
Traditional Stats	No	No	No	No	Yes	No	Yes
Sabermetric Stats	No	No	Yes	Yes	No	Yes	Yes
Team Fixed Effects	No	No	No	No	Yes	Yes	Yes
<i>N</i>	286	286	286	286	251	251	251
<i>R</i> ²	0.2407	0.3290	0.3924	0.3924	0.5832	0.5737	0.5968

t statistics in parentheses

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

Effect of International on ln(Salary) - Pitchers

	(1) ln_salary	(2) ln_salary	(3) ln_salary	(4) ln_salary	(5) ln_salary	(6) ln_salary	(7) ln_salary
war	0.0678** (3.27)	0.0631** (3.25)	0.0282 (0.79)	0.0282 (0.78)	0.00332 (0.06)	0.0212 (0.50)	0.0152 (0.25)
age	0.107*** (8.67)						
international	-0.181 (-1.82)	-0.172 (-1.84)	-0.0944 (-1.02)	-0.0942 (-1.01)	-0.0554 (-0.57)	-0.0297 (-0.31)	-0.0104 (-0.10)
year	-0.00353 (-0.19)	0.00454 (0.26)		-0.000630 (-0.03)			
servicetime		0.131*** (11.01)	0.133*** (11.13)	0.133*** (11.08)	0.122*** (9.91)	0.124*** (10.17)	0.124*** (9.70)
Constant	19.85 (0.54)	5.796 (0.17)	19.63*** (5.42)	20.91 (0.44)	14.35*** (11.58)	19.58*** (5.07)	18.92*** (4.54)
Traditional Stats	No	No	No	No	Yes	No	Yes
Sabermetric Stats	No	No	Yes	Yes	No	Yes	Yes
Team Fixed Effects	No	No	No	No	Yes	Yes	Yes
<i>N</i>	286	286	286	286	251	251	251
<i>R</i> ²	0.2456	0.3319	0.3940	0.3941	0.5832	0.5738	0.5968

t statistics in parentheses

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

Conclusion

Findings:

- There is statistically significant evidence to support that there is pay discrimination within the MLB for hitters
- ServiceTime is a better representative for 'experience'
- Traditional stats are still great variables to use when predicting player value

Next Steps:

- Further examine other factors on salary differentials
 - Player position, contract length, and free agency status
- Develop more accurate measure for a player's race
- Further examine the effectiveness of diversity and inclusion in sports