Labor Market Analysis and Evaluation: Senior Project

Department of Economics

Favoritism in Major League Soccer

Josh Fernandes

8 May 2017

Abstract

The purpose of this paper is to show if there is a relationship between where you are born and the wage you make in Major League Soccer. Using the Oaxaca Decomposition this paper will attempt to show that there is evidence of discrimination toward foreign born players in Major League Soccer versus their American born counterparts.

Table of Contents

Senior Project 2nd draft

I. Nature of the problem

Since Gary Becker first wrote his seminal work on the economics of discrimination many economists have looked into different labor markets to uncover ~~wonder~~ whether or not discrimination exists. Economists have been studying discrimination in the workplace for many years now and it seems that researchers cannot get a true measure of discrimination in the labor market. This might be because in the workplace the one variable you have to account for that is almost impossible to measure with accuracy is the worker's productivity. For example, you can try and make your argument that productivity is based on the number of hours an employee works at a firm for a given week, but this only measures the time spent at work not what was accomplished when working.

Productivity can be easily measured, however, in the sports field better than in any other sector of the economy. Sports gives you the opportunity to measure how a player produces on the field, court, or any playing surface you can think of. You can measure in basketball how many assists, points, and rebounds a single player has, or in football how many catches, yards, or turnovers a player has. Soccer is no different, there are many statistics that can help you measure the production of an individual player. Because of this, studying discrimination in sports can help widen the knowledge of discrimination in the workplace.

This opens up the questions that discrimination poses in sports such as if there is discrimination on player's wages in Major League Soccer. In soccer, there are players of many different ethnicities and many different races so measuring discrimination based on race or color makes the study very similar to others done in the past. Measuring discrimination based on country of origin makes the study unique. Not many people think of discrimination based on country of origin, mostly in the world today discrimination is based on skin color. Since soccer has such diversity among players and the study aims to look at a U.S. based league you can split up the groups into American born players verse foreign born players. Typically, previous studies have split players into blacks versus whites, but these studies have been done on the National Basketball Association or the National Football League where most of the players are from America. This study looks to improve upon the topic of discrimination in the workplace through adding new knowledge of what discrimination looks like in American sports, specifically Major League Soccer.

II. Literature review

In Burnett and Van Scyoc (2015), the author uses Gary Becker's theory of the three types of discrimination: Employer, Employee, Consumer/Customer. They argue that of the three types of discrimination, employer and customer are the two types they expect to find in the NFL for discrimination in the rookie wage market. The authors take a look at if there is discrimination in the NFL based on a sample of 333 rookie linebackers and 484 offensive linemen in the years 2001-2009. The authors use as a measure of the worker's performance as the ranking of the players based on their draft order and status. They cite the reason for not using in-game statistics as being hard to find proper statistics that measure performance for linebackers and lineman. Also since rookies are judged on performance of their rookie year not past statistics from when they played in college, they cannot use college statistics. They have 4 different equations, an OLS using the players wages as the dependent variable, a quantile regression earnings function, the Oaxaca decomposition, and a quantile treatment effects method. They take the results from the original OLS equations and use the Oaxaca decomposition to measure the wage difference due to discrimination. They come to the conclusion that there is no significant distinction between black and white players in the rookie wage market between offensive lineman and linebackers.

Holmes (2010) examines the baseball market for discrimination. For his data, he uses players that have just signed a new contract. He says that although this limits the data points he has, the data he uses is much stronger because players are paid for their performance rather than being paid and then performing. He makes up for the small number of people in the data set by adding additional years to his study. He uses quantile regression to see if there are differences in the lower half of salary ranges for black players. His argument for using quantile regression is that it has less restrictions than Least squares does and that Least squares gives you an average discrimination coefficient, but he wants to see if the coefficient might actually change for different percentiles of the data (Holmes 2010). Holmes finds that using quantile regression, there is a premium towards Whites and Hispanics of over 20% with respect to their black counterparts (Holmes 2010).

Pedace (2008) examines if the nationality affects the wage a certain player receives. His study is based in the English Premier league, thought to be the highest league in soccer. His initial empirical model comes from other papers but he focuses on wage differences by nationality. He expects there is a correlation between race and ethnicity with unobserved productivity characteristics which has been causing a biased estimate of the discrimination coefficients in the past. To fix this he uses a market test approach, an alternative form of measuring discrimination. Instead of running a model where log of wage equals productivity, team performance is the dependent variable. The author thinks this model counteracts the correlation between race and unobserved productivity characteristics. Pedace gets data from 1996-2002 from the *Soccerbase* to conduct his study. He starts off by using a random effects model then alters it slightly by adding in additional variables like attendance which he says may help identify fan preferences and increase team performance. His results suggest the players from South America receive preferential treatment in the wage market versus other players in the premiership. He attributes this to fans attendance actually increasing due to the more South American players on your roster.

In Mongeon (2015) the author proposes a new way to study discrimination in Hockey, using a market model, “A market model approach alleviates much of the concern relating to biased discrimination coefficients by regressing a team’s productivity (e.g., winning percentage) conditioned on a team’s payroll and race or ethnic-specific player inputs” (Mongeon 2015). The author proposes that since Kahn (1991) and Szymanski (2000), who talked about the effects of discrimination using earnings models, think earnings models lead to biased discrimination coefficients, and that the market model would better estimate the coefficient of discrimination. The author wishes run a test using the ethnicity of players who are French Canadian, American and international to the majority group, English Canadian as well as testing whether or not discrimination manifests itself in regions (Mongeon 2013). The author uses Weighted least squares regression, instrumental variables, and random effects regression models to estimate the model in his paper; wage equals other variables plus ~~the~~ a measure of discrimination. In his concluding remarks the author notes that there is evidence of discrimination found toward American and French Canadian players when playing for English Canadian teams.

Kuethe and Motamed (2010) discuss the way wages in the MLS are determined specifically for superstar players. They combine ~~pull~~ data on salaries of players from the MLS players union ~~and~~ with statistics from the MLS website. The model the authors uses is the natural log of the wage which is specified as a function of experience, performance, reputation, and a team success variable. Plus, they add in other variables to make sure they capture statistics of the players. They also control for where the player was from originally. They find that superstars are at a premium and therefore garner a significant wage difference from non-superstars in the league.

III. Data and Methodology

This paper will use an Oaxaca decomposition style approach to measure the difference in wage due to discrimination between foreign born players, players who are born outside the United States, and players born in the United States, in the MLS. The Oaxaca decomposition is a formula created by Ronald Oaxaca to measure discrimination. Given the dependent variable in this study is wages we know from Becker's theory that variables like experience and experience squared must be included in the model. Some other variables to be included in the model are age, the position the player plays (forward, midfield, defender, and goalkeeper), country of origin, and whether or not the player is a designated player. A designated player is one that earns a significant more than others on the team. This player would break the salary cap set by the League so the Other variables to describe players' productivity to be included are variables such as goals, assists, saves, shots, and a team performance indicator such as number of wins.

All of the statistics data such as goals, assists, and saves, comes from www.mlssoccer.com/stats. This website also has data on whether the player is designated player, where the player is from, and position the player plays. Wage data will come from www.mlsplayers.org/salary-info. Both sites have data stretching back to 2007 so that I can conduct my analysis over a ten-year period. This also means that some players will only have one year of data to look at while others will have possibly ten years of data.

My model includes variables like experience because that's what the theory calls for. The theory says experience affects the wage function in two distinct ways. One is as you get more experienced you should earn more, the other is that there comes a point in the professional career where your experience actually should start decreasing. These two measures of experience are shown in the model by experience and the square of experience.

This paper will use a semi-log model to show the effect of the variables on log of wages. Using a semi-log specification helps to control for skewness in the wage distribution which can be attributed to superstar effects (Lucifora & Simmons 2003). The model comes from a Mincer equation:

$$\ln(W_{it}) = \alpha_0 + \alpha_1 EXP_{it} + \alpha_2 PERF_{it} + \alpha_3 REP_{it} + \alpha_4 COUNTRY + \varepsilon_{it} ,$$

$Ln(W_{it})$ is the log of the wage of a player i in a given year t. EXP will be the experience and experience squared variables. PERF will be the set of performance statistics related to player i in a given year t. REP will be a dummy variable for the player's reputation in the league, whether he is a designated player or not where 1 will be the player is considered a designated player and 0 otherwise. And COUNTRY will be the dummy variable for the country of origin the player comes from, where 1 means that the player is from a country other than the United States and 0 means the player comes from the United States.

Each team has a maximum amount of money that they can spend on players: they will be fined if the total payroll exceeds that limit. However, Designated Players do not count towards their team's salary cap. Therefore, players with very high wages are tagged as a designated player by a team. These players from a soccer standpoint are seen as better players when they enter the league. For example, the last two Most Valuable Players, award given to the best player in the league as voted for by media, fans, and teams, in the league are designated players. This rule allows wages to be much higher for those special players which skews the wage distribution. But it also proves a point that these players need to live up to the wage they are being paid, which could motivate players to perform better than they normally do.

IV. Results

In the table 2 where log wage is dependent variable, the results came out different than I would have expected. The first results show no significant variables except for the coefficient on the Born1 dummy variable. This variable is significant at all levels and has a positive sign, which is hypothesized as the correct sign. The other variables in the model are not significant at any level so something is wrong with my SAS code. Also I only have 10 observations and you cannot run an OLS with less than 40 so that is probably the reason I am getting skewed results.

In the second OLS regression for goalkeepers, still log wage as the dependent variable, the results are a little more reasonable. Since everyone in the second regression is from the United States there is no coefficient on the dummy for Born1, because born1=0 for all of them. With 113 observations I am safe in running an OLS on this model.

The third OLS regression is for players in the field other than goalkeepers that are foreign born, this includes defenders, midfielders, and forwards. With the same dependent variable log wage, we see some strange results. 499 observations means I am definitely safe using OLS but my model is missing something and I don't know what it is yet.

The fourth regression using OLS is for players other than goalies that are born in the United States. Some of the other variables are significant and make sense though so I am on the right track at least a little bit.

V. Interpretation of results

The results in table 2 show the OLS results for goalkeepers that are foreign. As predicted in previous literature the results for the Designated player variable are positive and significant meaning that the "star" players in the league make more money than other players in the league. The experience variable is positive and significant meaning that as you increase experience in the MLS by one year your wage will increase by 8.5 percent.

The results in table 3 show the OLS results for goalkeepers that are from the United States. Once again the designated player variable is positive and significant but the coefficient is twice as great as the one in the foreign regression suggesting that goalkeepers born in the U.S. with star status have a greater wage than those born outside the U.S. Another result that is different from the previous results is the Goals Against Average variable, which is significant at the 5% level and positive. This result suggests that U.S. born goalkeepers make around 5% more for keeping goals out but their foreign counterparts do not have a significant coefficient.

The results in table 5 show the OLS for foreign players who are not goalies. These results show that goals and experience and are both significant and positive which implies that the more goals you score as a foreign player the higher your wage by a little over 6%, also the greater your experience the greater your wage.

The results in table 6 show the OLS for native players who are not goalies. Experience is positive and significant, more positive than for the foreign players meaning that the more experience a U.S based person has the greater their wage by almost 11% compared to their foreign-born counterparts.

Table 7 shows the Oaxaca decomposition for goalkeepers. The difference in skills is the explained portion of wage that is due to things like experience and saves. The difference due to discrimination is the part of the wage that cannot be explained by anything in the model. This is assumed to be discrimination because two people have the exact same skills but are paid different. The decomposition shows that around 50% of the wage can be attributed to differences in skills while around 50% can be attributed to favoritism towards foreign players.

Table 8 shows the Oaxaca decomposition for players who are not goalies. The decomposition here shows that the wage gap is around .42. The model explains about 95% of the wage gap meaning there is about 5% unexplained in the model. The model shows that about 28% of the wage gap can be explained by differences due to skills. While about 72% can be shown as favoritism towards foreign players.

VI. Conclusion and Limitations

Overall this study shows that there is a possibility of being born outside the US increasing your wage versus players born inside the US. One thing to consider is that the experience variable is only measured in years played in the MLS. Therefore, players that have played in leagues outside the MLS or started their careers outside the US and came to play here to finish have experience but not in MLS. Another thing to consider is the way that mlssoccer.com did not have any defending characteristics so defenders are not well represented in the model. This could also take some of the significance of the variables away meaning that coefficients could be misrepresented. This study does open up research into the field of MLS though. Hopefully this research will open up the possibility that more researchers can look into MLS players with regard towards discrimination. MLS players who come from overseas seem to be paid unfairly due to the fact that they are seen as a commodity. Not only can they bring in more revenue from shirt sales but they also attract fans to the games.

VII. Bibliography

Burnett, N. J., & Van Scyoc, L. J. (2015). Compensation discrimination for defensive players: Applying quantile regression to the national football league market for linebackers and offensive linemen. *Journal of Sports Economics, 16*(4), 375-389. Retrieved from http://search.proquest.com/docview/1728640970?accountid=14471

Holmes, P. (2011). New evidence of salary discrimination in major league baseball. *Labour Economics, 18*(3), 320-331. doi:http://dx.doi.org/10.1016/j.labeco.2010.11.009

Kuethe, T. H., & Motamed, M. (2010). Returns to stardom: Evidence from U.S. major league soccer. *Journal of Sports Economics, 11*(5), 567-579. Retrieved from http://search.proquest.com.ezproxy.uakron.edu:2048/docview/760190493?accountid=14471

Mongeon, K. (2015). A market test for ethnic discrimination in the national hockey league: A game-level panel data approach. *Journal of Sports Economics, 16*(5), 460-481. Retrieved from http://search.proquest.com/docview/1728641636?accountid=14471

Pedace, R. (2008). Earnings, performance, and nationality discrimination in a highly competitive labor market: An analysis of the english professional soccer league. *Journal of Sports Economics, 9*(2), 115-140. Retrieved from http://search.proquest.com/docview/56815392?accountid=14471

VIII. Appendices

Table 1: Descriptive Statistics for Goalkeepers

		country=0		country=1	
Variable	**Label**	**N**	**Mean**	**N**	**Mean**
Player_key		437	973.9199085	110	1039.85
Year		437	2011.84	110	2012.04
GP	**games played**	423	10.106383	110	12.5
GS	**games started**	423	9.9810875	110	12.3909091
MINS		423	898.5106383	110	1111.64
G	**goals**	423	0	110	0
SHTS	**shots**	423	0	110	0
SOG	**shots on goal**	423	0	110	0
GWG		423	0	110	0
HomeG		423	0	110	0
RoadG		423	0	110	0
G_90min		423	0	110	0
SC_		423	0	110	0
SHTSF		423	46.356974	110	55.6
SV	**saves**	423	31.8037825	110	38.7090909
GA	**goals against**	423	13.5650118	110	15.8272727
GAA	**goals against average**	423	0.9763593	110	0.9485455
W	**wins**	423	3.5910165	110	4.7181818
L	**losses**	423	3.7068558	110	4.2363636
T	**ties**	423	2.6903073	110	3.3818182
ShO	**number of shutouts**	423	2.5673759	110	3.5363636
W_	**win percentage**	423	20.3919622	110	24.1190909

Sv_	**save percentage**	423	44.4262411	110	47.5936364
Exp	**experience**	423	3.2458629	110	2.5454545
VAR8	**wage**	437	83435.82	110	110035.55
Comp		437	89445.07	110	121827.35
lnwage		436	11.0347121	110	11.3467548
DP1	**designated players**	437	0.0343249	110	0.0090909

Table 2: Foreign Goalkeepers

Variable	**Label**	**DF**	**Parameter Estimate**	**t Value**
Intercept	Intercept	**1**	10.6578	103.24
GS	games started	**1**	0.06527	4.92
Exp	experience	**1**	0.08506	3.1
SV	saves	**1**	-0.01069	-2.55
GAA	goals against average	**1**	-0.05496	-0.76
W_	win percentage	**1**	0.00464	1.69
DP1	designated players	**1**	1.92725	4.07
			R Squared	0.61

Table 3: Native Goalkeepers

Variable	**Label**	**DF**	**Parameter Estimate**	**t Value**
Intercept	Intercept	**1**	10.25698	238.98
GS	games started	**1**	0.02352	2.67
Exp	experience	**1**	0.13852	12.78
SV	saves	**1**	0.0015	0.57
GAA	goals against average	**1**	0.04779	1.83
W_	win percentage	**1**	0.00024255	0.22
DP1	designated players	**1**	3.6192	8.1
			R Squared	0.62

Table 4: Descriptive Statistics for Field Players					
		country=0		country=1	
Variable	**Label**	**N**	**Mea**	**N**	**Mean**
GP	**games played**	2306	15.7	2023	16.8314
GS	**games started**	2306	12.5	2023	13.0796
MINS		2306	1125	2023	1168.15
G	**goals**	2306	1.32	2023	2.08848
SHTS	**shots**	2306	12.7	2023	19.1696
SOG	**shots on goal**	2306	4.73	2023	7.174
GWG		2306	0.37	2023	0.5561
HomeG		2306	0.76	2023	1.25507
RoadG		2306	0.55	2023	0.83342
G_90min		2306	0.11	2023	0.13671
Exp	**experienc**	2306	3.18	2023	2.37667
VAR8	**wage**	2359	####	2023	253943
Comp		2359	####	2023	284351
lnwage		2359	11.2	2023	11.6326
DP1	**designate**	2359	0.06	2023	0.14533

Table 5: Foreign Field				
Intercept	Intercept	**1**	10.9513	323.26
Exp	experience	**1**	0.04841	4.77
GS	games started	**1**	0.02597	12.32
G	goals	**1**	0.06096	4.74
SOG	shots on goal	**1**	-0.0125	-2.64
DP1	designated players	**1**	1.2984	23.7
			r squared	0.4197

Table 6: Native Field				
Intercept	Intercept	**1**	10.3447	441.01
Exp	experience	**1**	0.15158	25.77
GS	games started	**1**	0.02017	14.55
G	goals	**1**	0.01599	1.45
SOG	shots on goal	**1**	0.01236	3.06
DP1	designated players	**1**	1.29476	18.8
			r squared	0.515

Table 7. Difference in endowments for Goalkeepers.

Variables	Coefficients		Means		Difference in Skills	Possible Discrimination
	White	Black	White	Black		
Intercepts	10.25698	10.6578				0.40082
Games started	0.02352	0.06527	9.9810875	12.390909	0.16	0.42
EXP	0.13852	0.08506	3.2458629	2.5454545	(0.06)	(0.17)
Saves	0.0015	-0.01069	31.803783	38.709091	(0.074)	(0.39)
Goal against avg	0.04779	-0.05496	0.9763593	0.9485455	0.00	(0.10)
Win %	0.00024255	0.00464	20.391962	24.119091	0.02	0.09
Designated player	3.6192	1.92725	0.0343249	0.0090909	(0.05)	-0.058076015
					(0.01)	0.19
	Wage Gap	0.312	Total Difference in Skills		0.182	
	differ	0.58222491	discrimination	0.6011811		1.183405984

Table 8. Difference in endowments for Field Players.

Variables	Coefficients		Means		Difference in Skills	Possible Discrimination
	White	Black	White	Black		
Intercepts	10.95131	10.34465				0.60666
Games Started	0.02597	0.02017	13.079585	12.512576	0.01	0.07
EXP	0.04841	0.15158	2.3766683	3.179098	(0.04)	(0.33)
goals	0.06096	0.01599	2.0884825	1.3191674	0.05	0.06
shots on goal	-0.01247	0.01236	7.173999	4.725065	(0.03)	(0.12)
Designated player	1.2984	1.29476	0.1453287	0.0584994	0.11	0.00
					0.10	0.29
	11.632583	11.2064591				
	Wage Gap	0.426				
			Total Difference in Skills		0.398	
	differ	0.25	ence due to discrimi	0.6886681		0.94