An Exploration into the Relationship of MLB Player Salary and Performance

Statement of the Problem:
When considering professionals sports leagues, the common topic of discussion is usually about individual players or favorite hometown teams. Sports enthusiasts are the first to rattle off a laundry list of statistics about teams and players, but do they realize they are openly discussing that player’s efficiency in terms of labor? Although it is a popular form of entertainment there is a market structure to our professional sports leagues that is comparable to a perfectly competitive labor market. With so much attention paid to sports today it would be fitting to discuss America’s past time: Baseball.

Principal Hypothesis with Motivation:
The MLB is the premiere baseball league in the world and features the top global talent. Pair this with the fact there is copious amounts of data that depicts player’s performances, therefore, the MLB provides great insight when exploring the relationship between employer and employee. As of November, 2014 the lowest MLB salary settled at $507,500 per year with the highest being around $32,000,000 per year (spotrac.com).

As in most firms, the productivity of the employees generates revenues for the employer. Baseball can be represented similarly if you consider the team the employer and the player’s employees. The productivity created by these employees is a game of baseball. MLB team owners attempt to get the most productive athletes at the lowest price, but unlike the majority of markets the employees have a degree of bargaining power over their wages. Referred to as the “contract zone’, players and team owners agree what wages should be paid. The lowest productivity parameter is known as the “replacement level”, which states that replacement player’s salary and performance is only marginally above that of the top player not contracted to play in the league. Meaning, the best available free agent MLB player becomes the replacement player once hired by a team.

There is also another economic factor presented within the MLB, which is player mobility which play a part in the valuation of a salary. In turn, the team is hoping to increase its wins while still maintaining profit margins. How do MLB teams determine valuation of players? For this paper, we will test if an individual player’s salary is determined by their on-field performance.
Null Hypothesis: An MLB player’s performance has no impact on salary

Alternative Hypothesis: An MLB player’s performance has a positive impact on salary

Review of the Literature:

To date, there has been a considerable amount of literature that explores any causality between player salary and performance in the MLB. As prior research has been done to test for causality between individual player salaries and performance. To properly test for a link in salary and performance, we need to understand some of the intuition how wages are determined in the MLB.

The findings from a popular book by Scully (1989) concludes that team revenues are directly related to the teams win percentage. Higher team revenues give owners the ability to pay higher wages to players. As the old adage goes, you get what you pay for. These results were reinforced in Hasan (2008) in which she states cases of small market team success are outliers and team success is highly influenced big payrolls. This is accomplished by using the equation: \( \text{Win Pct} = \alpha + \beta \text{Pay-scalet} + \varepsilon_t \), a relationship was examined between pay and performance in both the team and the overall pooled level.

Hall, Szymanski and Zimbalist (2002) used a 20 year data set to examine this link in both MLB and English Soccer. Additionally, they also tested for causality in an attempt to explain whether the relationship runs from payroll to performance or the reciprocal. In this paper, the authors were able to demonstrate that the relationship between performance and payroll increased significantly during the 1990s as compared to the 1980s, however, they failed to establish the direction of causality.

In a related study, Burger and Walters (2003) proved the existence of market size effect on expected team performance. Backed by the regression results, they focus on the interaction between a team’s win percent and the market size (population) of its home town. Then, using their estimated revenue function to compute the marginal revenue of a win for all 30 MLB teams they derived that the value of an additional win for a New York team ($3.62 million) is approximately six times the value of an additional win in Milwaukee ($0.59 million). This allows for some teams to pay more for a player with a performance level that is equivalent to players in a smaller salary bracket.

Furthermore, when examining Depken II and Wilson (2004) a comprehensive breakdown of an MLB players marginal revenue product (MRP) which explains the change in total revenue for a ball club when adding another labor unit, in this case another player to a team. The model used in this research allows for the measurement of individual player MRP and compare it to actual salary. However, this method may be effective but it is not practical when looking at a larger sample size of players.

The panel approach was used in the Hakes and Turner (2009), which points out there is a systematic bias in the regression coefficients when pooling all players into the same
“performance” category if player’s value is placed on the player’s talent. This was counteracted by dividing players into talent quintiles. The crux of their work mainly focused on examining trends in player productivity and salaries as player’s age. With this, they were able to provide support for evidence of causality that free agent players are paid proportionately to production value.

**Formulation of the Model:**

To formulate the model, we want to focus on our hypothesis and how to best investigate a relationship between performance and salary. As confirmed by the literature review, we have some parameters had to be set which excluded any playoff games, which means we are using a standard 162 games season. Involving the post season includes extra incentives which would not properly explain a causality between performance and salary. This would also severely limit our field of testing.

**Why Pitchers are excluded from the model:**

Pitchers were also removed from the data because of the statistics used to evaluate pitcher performance varies from the rest of the players. Pitchers are to dependent on their team in terms of run support and reliance on the rest of a bullpen. For instance, a pitcher can gain a win only after his teams has scored runs and a relief pitcher closes the game. There is also a large disparity between inning pitched between starting and closing pitchers.

**The Data:**

Generally, when evaluating a player we consider the rudimentary statics such as homeruns, runs batted in (RBI), batting average, and run. However, many factors may contradict that the actual player’s performance gives these statistics any significance. For example, homeruns may be affected by the dimensions of the stadium, the weather or even the altitude. RBI’s and runs tend to be based on a teammate’s performance and batting average can just be luck. These factors are not an accurate portrayal of a players performance, thus we must derive other ways to measure it. When evaluating players performance or productivity the first parameter in this model is OPS (on base percentage plus slugging), which has been shown to be both simple to calculate and an accurate predictor of team output (wins). As OPS measures production per unit of playing time. Experience is time variables that represent the depreciation of labor. Productivity is a function of performance output and depreciation of the athlete.

Our base equation will be \( \text{Salary} = \beta_0 + \beta_1 \text{OPS} + \beta_2 \text{Exp} + \beta_3 \text{SB} + \beta_4 \text{market\_size} + e_1 \)

The dependent variable in question is salary. The test will confirm if there is any statistical significance of “On base percentage plus slugging” (OPS), “Experience” (exp), “stolen bases” (SB) & “market size (market\_size) to player’s salary.
**Salary:** Salary will be defined as the gross total contractually agreed to pay a player over the course of a given year. Available data on players salary dates back at least 25 years, however for our model we used a panel data set of 42 players salaries from 2009-2014.

**SB:** Stolen bases is another offensive statistic that relies solely on the individual player. As an exciting and pivotal function in the game of baseball, I believe stolen bases will positively relate to a players salary.

**Exp:** Experience will serve as a non-linear regressor, thus it will be squared. Experience is measured as years spent in the MLB. The parameter estimate for “exp” should be stated as positive. Intuitively the logic is the longer you have been in the league, the greater level of experience a player would have.

**OPS:** On base percentage plus slugging is referred to as a “sabermetric” baseball statistic that calculates the sum of a player’s on base percentage and slugging average. We use this measure instead of a standard on base percentage (OBP) because the OBP over values walks. The slugging percentage is a function of total bases divided by at bats which is combined with OBP to achieve a better fitting model. The following is the formula used to derive that OPS function that was created within the data set. I expect the parameter estimate to be positive as this sums up offensive performance.

\[
OPS = \frac{AB \ast (H + BB + HBP) + TB \ast (AB + BB + SF + HBP)}{AB \ast (AB + BB + SF + HBP)}
\]

**Data Sources & Description:**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary</td>
<td>Gross annual wages paid to each player from 1995-2010</td>
<td>Rankings (Spotrac.com) <a href="http://www.spotrac.com/rankings/mlb/">http://www.spotrac.com/rankings/mlb/</a></td>
</tr>
<tr>
<td>OPS</td>
<td>On Base Percentage plus Slugging is a full measure of offensive productivity from 1995-2010</td>
<td>Lahman’s Baseball Database (Sean Lahman) <a href="http://www.seanlahman.com/baseball-archive/statistics/">http://www.seanlahman.com/baseball-archive/statistics/</a></td>
</tr>
</tbody>
</table>
Exp

Experience is defined at number of years in the MLB Lahman’s Baseball Database (Sean Lahman) http://www.seanlahman.com/baseball-archive/statistics/

Market_size

Total market value of an MLB team that determines payroll (in millions of dollars) http://deadspin.com/2014-payrolls-and-salaries-for-every-mlb-team-1551868969

**Model Estimation & Hypothesis Testing:**

**TABLE 1**

The Effect of an MLB Player’s Performance on Salary  
(Independent variable: Salary)

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-12639178</td>
<td>(-2.79)</td>
</tr>
<tr>
<td>On Base Percentage Plus Slugging [ops]</td>
<td>121809**</td>
<td>(2.44)</td>
</tr>
<tr>
<td>Market Size [market_size]</td>
<td>0.04964**</td>
<td>(2.02)</td>
</tr>
<tr>
<td>Stolen bases [SB]</td>
<td>-198482</td>
<td>(-1.57)</td>
</tr>
<tr>
<td>Experience [exp]</td>
<td>.0015</td>
<td>(1.11)</td>
</tr>
</tbody>
</table>

R-Squared | .4915
Adjusted R-Squared | .4365
Number of Observations | 42
---|---
F- Value | 8.94

Note: The symbols ** and *, respectively, denote statistical significance at the 5% (or better) and 10% levels.

**Interpretation of the Results:**

Running the regression has posed some interesting results. My adjusted $r^2$ value of .4365 explains about 43% of the variance within the model. This value shows the model displays some degree of explanatory power, however, this level is not very strong. This adjusted $r^2$ value does suggest that with some modifications to the correct variables, that this model may hold a higher degree of explanation for the hypothesis test.

In regards to variables, it seems that stolen bases (SB) did not produce any promising results as it yield a negative parameter estimate of -198482 at a t-value of -1.57 that states no significance. This parameter estimate which would suggest that each stolen base would decrease salary.

A similar story can be told about experience (exp). The parameter estimate of .0015 and a t-value of 1.11 show absolutely no significance. This suggests there could be problems with multicollinearity or the variable was not properly normalized into the data.

Conversely, the market size (market_size) and on base percentage plus slugging (ops) displayed some faith in the model. Reviewing market size, we see a positive parameter estimate of .04964 which states that market size could account for up to about 5% of a salary increase for a player. At a t-value of 2.02 this is statistically significant at the 99% level.

Finally, OPS was able to represent a player’s offensive performance and does suggest some causality of player salary increase. As the signs returned were positive, we conclude a t-value of 2.44 and a highly statistical significance. This suggests that for every increase in an OPS points that a player’s salary could see an increase up to $121,809. As the prior literature claimed, the OPS calculation provides a basis of performance.

**Limitations of the Study:**

After careful consideration and viewing the results of the regression there is room to improve some of the errors in the model. For example, I could simply add a “spec white” test to conclude any heteroskedasticity between the variable. The difficulty was finding variables that would properly represent the data for MLB players. Batting statics are the only measure of offensive output, but it does not account for players that are excellent defensive players which may be highly correlated to salary as well. The OPS calculation in the data was carried out with precision but may not be the most accurate way to sum up offensive value.
The stolen base variable, which was added as an attempt to help the model did just the opposite. It proved to be poorly related to salary and may only represent a specific skillset of a player.

Experience is another variable under investigation. I have no reasonable explanation as to why this variable did not cooperate with the data. One solution may be to use an age variable instead of experience. Again, if looking at this in terms of labor then age accounts for depreciation of labor capital and may reflect a better determination of salary. This idea stems from the literature review.

A large problem within this study as well is that it is difficult to allocate salaries across the differences of baseball positions. Excluding pitchers makes sense when just including offensive statistics, but there is no account for defensive value within this model. One major change that could complicate the model, yet benefit it would be a separate test ran on defensive variables along with the current model.

**Conclusions:**

After conducting this study, it is safe to say this model does hold some power in regards to explaining MLB player’s salary but this is a mild attempt to assist the alternative hypothesis at best. In order to have a more effective measurement, the adjusted r-squared value needs to increase and a dummy variable needs to be inserted to account for stolen bases that bared ineffective.

To sum up the hypothesis test, we cannot reject the null hypothesis that a player’s performance has no effect on salary. While we recognize that market size has an effect on player’s salary. This suggests that teams with a larger payroll with over pay for players of a similar performance level simply because they can afford too. However, this is not the case for smaller market teams so we see a little bit of income disparity depending on the team. The OPS shows some relationship to determining a player’s salary. Players with high OPS tend to be the league best producing players and even at a quick glance of tabled data, there appears to be a relationship. Overall, there is not enough evidence that there is causality between the two. The values were just too weak to support the original hypothesis.

**References:**

Baseball Reference (Baseball-Reference.com)
http://www.baseball-reference.com


Wilson, Dennis P. "Labor Markets in the Classroom: Marginal Revenue Product in Major League Baseball."

Data:

Lahman’s Baseball Database (Sean Lahman) http://www.seanlahman.com/baseball-archive/statistics/

Rankings (Spotrac.com) http://www.spotrac.com/rankings/mlb/