MIT Sloan Sports Analytics Conference 2011
March 4-5, 2011, Boston, MA, USA

Workforce Analytics in Baseball Player Management

Michael Greene, Adam Hirsch, ACAS, MAAA
Deloitte Consulting LLP
Boston, MA, USA 02116
Los Angeles, CA, USA 90071
Email: migreene@deloitte.com, ahirsch@deloitte.com

Abstract

The "Sabermetric Revolution" has brought rigorous analytics to the fore in operating and managing baseball teams. In sports, Baseball is unique in the data that are available: the isolated pitcher-versus-hitter moments allow the game to be effectively analyzed through the statistics of individuals. Over time, front offices have increasingly relied on developing these statistics and sought to maximize expected performance per dollar spent. However, the over-abundance of data and well established metrics has prevented organizations from most effectively utilizing analytics to manage their primary workforce: the players. This paper will provide a framework for the application of statistical tools to roster management and development, allowing a team to leverage the power of analytics and effectively incorporate Sabermetric tools and techniques to improve on-field performance. We will first analyze performance through the lens of Workforce Analytics, a statistical modeling framework which can be applied to any industry. Next we will define an appropriate variable to incorporate future value and cost: Expected Performance Efficiency. Lastly, we will provide a case study of Workforce Analytics as applied to the World Champion San Francisco Giants entering the 2011 season.

Introduction

Despite the available tools to study baseball player values, a team's success is not simply the sum of twenty five random variables. World Series teams are seldom comprised of many of the same "type" of baseball player. Rather there is a science to creating the optimal team, in addition to the optimal players. As statistical measures improve, how can teams work within roster and salary constraints to construct the optimal team? How can a front office use career forecasts and expected future value to produce a team with the maximum expected output?

The answer is Workforce Analytics, used in other industries to leverage predictive models and optimization techniques to isolate and improve key performance indicators in a world where ever expanding databases collect information on the workforce. Using Workforce Analytics, companies in many industries have begun effectively using data and statistical tools to improve productivity and efficiency, demonstrating the value of effectively applied analytics.

Properly applied, the Workforce Analytics framework can empower organizations to answer complex questions and assemble the optimal baseball roster. Workforce Analytics is not the derivation of new metrics, but a rigorous conglomeration of existing Sabermetric measurements, planning, statistical forecasting, and optimization techniques, which can maximize expected future wins within business constraints, such as those imposed by contracts, salary, and available talent.

When Data Becomes Too Much Data

In the Sabermetrics era, General Managers do not have to look hard to find data. Employees and even fans are analyzing players in a myriad of ways using countless variations of commonplace statistics. At first glance, this wealth of data appears to be a blessing, offering a competitive advantage and a unique lens through which one can assess performance. However, the sheer magnitude of data can reach a point of diminishing marginal returns. Unless the available information is properly utilized, it becomes a potential risk: misinformed decisions can be costly. This problem is not unique to baseball, as managers and executives in many industries have access to reams of data, yet few tools to use that data effectively to improve the business.

Properly utilizing the available data in a meaningful way is the catalyst for a statistical roster optimization. Front office personnel must move beyond the analysis of statistical measures to put models into action in a systematic way and integrate different model content to construct an exhaustive business approach.

Baseball professionals and fanatics already analyze performance, calculate injury risk and forecast career progression, and track contract negotiations. But to translate these numbers into business decisions, we must bridge the gaps between academic studies and the business operation.

Workforce Analytics can be utilized to manage the roster, employing a top-down approach to analyzing the workforce to fuse available data and operational goals. Applying predictive models to the team roster can synthesize the existing understanding of players with the innovative recommendations of powerful models. Ultimately the foundation for this methodology is as simple as the topic of many introductory economics courses - Supply and Demand - and some of the unique aspects of the baseball industry provide an even greater opportunity for this fusion of analytical tools.

A Diamond in the Rough

In this context, Supply is the pool of potential baseball players at a position. General Managers can call upon both their minor league systems and the available free agent market to fill slots on the 25-man rosters. However, they must consider the current contracts in place or contract demands, in value and amount. A marquee free agent may be available and fit the positional needs of a given team, but if contract demands don't mesh with the strategy of the team (either the salary or the length), the supply is further limited. Several methods of player evaluation already include this concept of Supply by comparing a player to the hypothetical "replacement player." We extend this concept to compare not only to the hypothetical replacement, but to compare to specific players at known or forecasted contracts. Although impossible to do manually, optimization algorithms allow for exploring the astronomical permutations to identify the optimal selection.

Demand is represented by the positions General Managers have to fill. For most of the season, there are 25 spots on a roster and 9 positions on the baseball field, but teams typically have many ongoing contracts at any one point in time. Therefore, any planning analysis must consider the length of player contracts and incorporate risks and variability that can impact demand and therefore value:

1. Injury risk – As players suffer from injuries, they must be replaced with other personnel, either short term or long term.

2. Attrition risk – Some players decide to leave the club as their contracts expire, despite interest from the organization of resigning them.

3. Performance risk - An additional consideration that encompasses several risks related to the performance of the players under contract at the beginning of a season.

Thus, this framework begins with a simple question of Supply and Demand: which roster spots need to be filled and which players are available to fill those spots at what value and cost to the team?

Although the team roster can be viewed as 25 puzzles, instead a team must be viewed holistically: as a decision impacting a greater business strategy. To accomplish this, we define **Expected Performance Efficiency (EPE)**, the expected value that a player will deliver per dollar of salary. For each position, we will define the metric EPE representing our performance value output as:

*Expected Performance Efficiency = [Expected Performance Value] * [Expected % of Season] / Salary*

This formula does not replace current value metrics, but allows them to be used in a prospective manner. Expected Performance Value relies on a measure of contribution, for which an abundance of metrics exist. The Expected Percent of Season is the amount of the season that a player is expected to contribute. Here, we attempt to divorce the player's estimated future health from the player's potential performance. Injury risks are different from position to position and player to player, but can be forecasted depending on age and past injury history. This expected performance is then divided by salary, creating a metric where we can compare players based on value per dollar spent.

For our case study, we have chosen Wins Above Replacement ("WAR") as the value metric. WAR measures the number of wins a player is expected to contribute above the expected contribution of a replacement player at the same position with the same opportunity (equal number of plate appearances). WAR does include defensive contributions, and therefore is a robust metric to evaluate the Supply and Demand of players – the statistic uses as its benchmark a fictitious player who performs at the "replacement level," a level of performance that can be expected at minimal cost.

However, WAR is not an estimate of the total value contributed to the team, rather it measures on-field production and excludes any consideration of leadership and clubhouse chemistry. Personality traits and clubhouse characteristics can affect value. In this paper, we use WAR alone. Future research into the value of less tangible traits can be included in the Workforce Analytics framework within the EPE calculation.

Within the EPE calculation, we estimate future WAR based on prorating each player's WAR from 2010 and extrapolating to 600 plate appearances. For each player, we divided the statistic by plate appearances in 2010, and then multiplied by 600. This will give us a uniform statistic, independent of health or plate appearances – the value delivered if he played the whole season. Finally, we must estimate the amount of time the player will see within the team construct. For the case study, we analyzed each position for the Giants and forecast the amount of time each player would see at the position, weighted by age and historic injury history. This heuristic measurement represents the *Expected % of Season* in the EPE formula.

For salary, we will consider the guaranteed portion (including signing bonus) of a roster player's contract for the current year and the expected market value for a free agent. An enhancement to the EPE formula would be to incorporate future value by including the length of contract, expected value of the next contract, and projections of future years' performance. A potential improvement would be a multiple year view: weighting future years using a Net Present Value calculation would allow teams to take a longer view in planning team performance.

To best leverage all these components into one methodology, the ideal case analysis would involve the following components:

1. Inputs – the data flowing into the statistical model will thus consist of the following for every major league player and the average minor league player by age:
    a. WAR to measure *Expected Performance Value*
    b. Injury tiers and positional hierarchy to measure *Expected % of Season*
    c. Guaranteed salary for the season to measure *Salary*
2. Optimization Engine – utilize optimization algorithms to maximize EPE within boundaries and constraints, such as total payroll and existing contracts.
3. Simulation Engine – rather than producing a point estimate, incorporate variance in outcomes through simulation.
4. Decision – the model only produces indications and suggestions, it is not a replacement for common sense.

Case Study: The 2010 San Francisco Giants

To illustrate the concept of applying a simplified Workforce Analytics framework to a team, we analyzed the World Champion San Francisco Giants entering the 2011 baseball season. The Giants won the 2010 World Series on the strength of their pitching. But can we dig deeper to better understand their strengths and weaknesses as they enter the 2011 season? To go beyond the normal roster analysis, we apply Workforce Analytics, casting the potential roster options in terms of Supply and Demand and facilitating a clearer picture of potential roster moves the Giants could make to improve the team on the field. We do not use specific predictive models to forecast WAR, time played or injuries, nor do we employ optimization techniques. This case study is intended to demonstrate the use of analytics and models to measure efficiency. More accurate and detailed forecasts improve this analysis by providing new components, rather than replacing the framework.

Although the Giants are strong and efficient in the pitching department, they also possess an infield that is cost-efficient – relatively inexpensive for the on-field value. For each player on the Giants' roster, we obtained the 2011 salary, contract, and player value statistics [1]. As an estimate of the value for 2011, we assumed that each player will have the same rate of WAR per plate appearance as seen in 2010. Finally, we forecast the amount of playing time that each player will see next year. This simple metric is based on how we expect the Giants to allocate playing time at each position. As a result, we can analyze the Demand – which positions the Giants have available, and the Supply – which players they are currently filling in each position. Our analysis indicates that the Giants have assembled an efficient team in the infield and on the mound, but have spent additional money for little expected return in the outfield. From an efficiency standpoint, a potential source of improvement would be better investment in the outfield, while trading some of their expensive players who provide little value on field.

Infield

In the infield, the Giants have assembled an efficient group – players that deliver high contributions to winning without high cost. Viewed slightly differently, the majority of the Giants infield is paid an amount that is justified by their contribution value. For example, the player that delivered the most value in 2010 was Aubrey Huff. Huff is scheduled to make $11m in 2011, but brings value, as he delivered the highest WAR for the team in 2010. With an expected WAR of 4.77 in 2011 (including *Expected % of Season*), Huff will likely provide almost 5 wins in the standings over a replacement player. Taking this a step farther, we can calculate Huff's EPE, at 0.4336 expected wins per million dollars of salary.

For a comparison to other players, see Chart 1 below. This chart shows the expected WAR on the x-axis and the corresponding salary on the y-axis. A linear trend regression line is fit to the data. Notice the upward slope, with a parameter estimate of $2.0m/WAR. This implies that, on average, the Giants paid about $2m per win over replacement from players in the infield.

A full account of EPE for the infield can be found in Table 1 below. Note that Huff does not have the highest EPE in the Giants infield. Pablo Sandoval is the most efficient player as measured by EPE, at 1.67 wins above replacement per million dollars. Sandoval's relative value is so high because he is contributing significantly while still being paid a minimal salary. Clearly, players with negative expected WAR are difficult to justify, but Bengie Molina stands out, being paid $4.5m while delivering average catching value at this stage in his career. Miguel Tejada, an offseason addition to replace Edgar Renteria, actually provides a less efficient alternative than Renteria. While costing the Giants $6m next year, Tejada is expected to generate less than 0.5 wins above replacement, while Renteria was expected to provide 1.7 wins and is expected to sign with the Reds for less than $3m [2]. As a final note, we have included Buster Posey's $6m signing bonus, which is a one-year cost [3]. Given the expected benefits from this talented young catcher, this single year EPE may be taken with a grain of salt. If a multiyear view were calculated, he would appear as a more efficient source of value.

Outfield

The Giants' outfield is a different story in terms of efficiency. In fact, viewing the entire outfield in terms of EPE in Chart 2, one will notice that there is not a clear pattern of paying for value. Andres Torres is expected to deliver 3.8 wins over replacement in 2011 while getting paid only $426,000. The low points in efficiency, however, are due to Jose Guillen and Aaron Rowand, paid $12m and $13.6m respectively while delivering average production. A full table of EPE for the outfield can be found in the Appendix as Table 2.

Pitching

The Giants' vaunted pitching staff is remarkable not only in their contribution to the recent World Series title, but also with respect to efficiency. The comparison of expected value and salary is shown in Chart 3 below. Notice that the Giants pay only $1m per win over replacement, significantly less than that for their infield. One particular pitcher stands out as an outlier: Barry Zito. His salary of $18.5m is the highest on the Giants' roster, and he is expected to add only one win over a replacement player. Excluding Zito from the regression does not affect the estimate, since his WAR is close to the team mean. The remarkable individuals on the Giants' staff are Casilla, Romo, Sanchez, and Bumgarner. These players deliver high value, roughly 9.5 wins, while being paid less than $6m in total. The full calculation of EPE is given in the appendix in Table 4. Note that for Bumgarner, similar to Posey, we have included his signing bonus [4].

Conclusions

The application of Workforce Analytics to baseball roster management is intended to build upon existing statistical measures, rather than replace any such metrics. This methodology can synthesize research to assist General Managers in making efficient and effective business decisions using analytical support.

We have built this study upon several assumptions, and we would like to mention a few potential improvements going forward.

1. Data inputs – potential improvements include but are not limited to:
    a. Expected Performance Value – develop a statistic that includes the value of WAR but also considers a player's intangible value, such as leadership qualities and chemistry
    b. Expected % of Season – develop a metric that incorporates actuarial models and more rigorous analysis of injury and health data
    c. Salary – enhance the formula by incorporating length of contract and expected next contract
2. Retrospective analysis – apply this methodology to the Giants and other teams' rosters one year ago to see how this technique would have helped for the 2010 season

We welcome any related discussion or questions.

References

[1] "2010 San Francisco Giants Batting, Pitching, and Fielding Statistics." *Baseball-Reference.com.*
<http://www.baseball-reference.com/teams/SFG/2010.shtml>

[2] "Reds to Sign Edgard Renteria." *Mlbtraderumors.com.*
<http://www.mlbtraderumors.com/2011/01/reds-to-sign-edgar-renteria.html>

[3] "Giants Recall Buster Posey." *Mlbtraderumors.com.*
<http://www.mlbtraderumors.com/2010/05/giants-recall-buster-posey.html>

[4] "Madison Bumgarner." *Thebaseballcube.com.*
<http://www.thebaseballcube.com/transactions/B/Madison-Bumgarner.shtml>

Appendices

Chart 1: Infield Efficiency

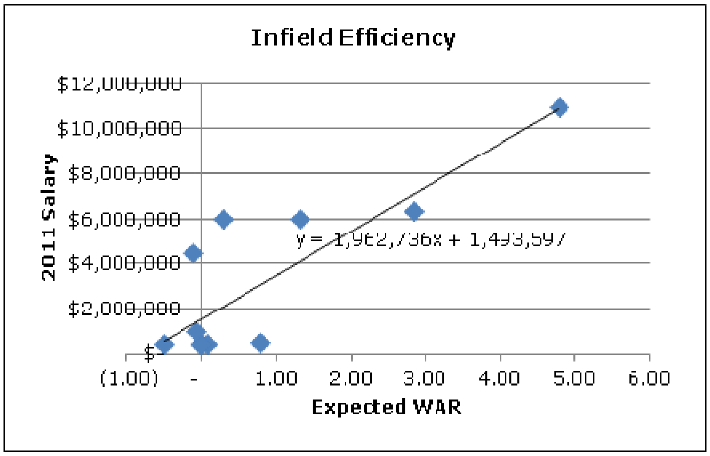

Table 1: Infield Efficiency

Player	Position	Age	Expected % Season Played	Expected WAR	Expected Salary + Bonus	Expected Performance Efficiency (per $1m)
Pablo Sandoval	3B	23	100%	0.78	$ 465,000	1.6757
Aubrey Huff	1B	33	90%	4.77	$ 11,000,000	0.4336
Buster Posey	C	23	70%	2.84	$ 6,400,000	0.4444
Freddy Sanchez	2B	32	70%	1.32	$ 6,000,000	0.2192
Matt Downs	2B	26	10%	0.07	$ 400,000	0.1705
Miguel Tejada	SS	36	75%	0.28	$ 6,000,000	0.0460
John Bowker	PH	26	20%	-	$ 410,000	-
Emmanuel Burriss	2B	25	30%	-	$ 410,000	-
Travis Ishikawa	1B	26	10%	-	$ 417,000	-
Eli Whiteside	C	30	15%	-	$ 405,000	-
Bengie Molina	C	35	15%	(0.12)	$ 4,500,000	(0.0271)
Mike Fontenot	2B	30	15%	(0.06)	$ 1,000,000	(0.0602)
Ryan Rohlinger	PH	26	5%	(0.50)	$ 400,000	(1.2500)

Chart 2: Outfield Efficiency

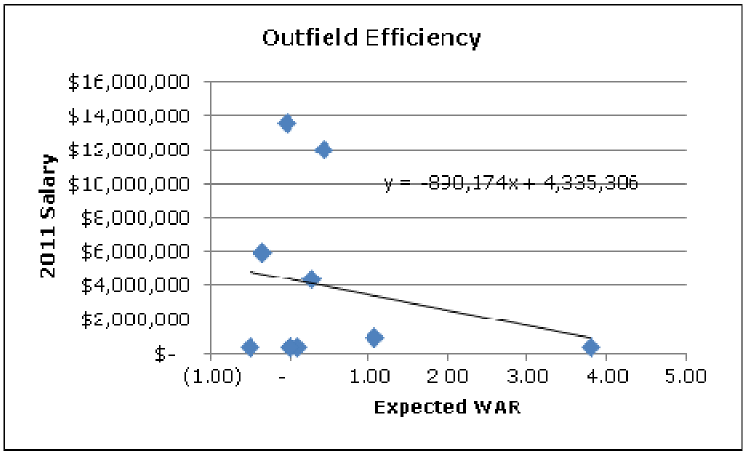

Table 2: Outfield Efficiency

Player	Position	Age	Expected % Season Played	Expected WAR	Expected Salary + Bonus	Expected Performance Efficiency (per $1m)
Andres Torres	CF	32	80%	3.79	$ 426,000	8.8955
Pat Burrell	LF	33	20%	1.06	$ 1,000,000	1.0557
Nate Schierholtz	RF	26	20%	0.10	$ 416,500	0.2287
Cody Ross	LF	29	60%	0.27	$ 4,450,000	0.0610
Jose Guillen	RF	34	80%	0.43	$ 12,000,000	0.0354
Eugenio Velez	LF	28	10%	-	$ 416,000	-
John Bowker	PH	26	20%	-	$ 410,000	-
Aaron Rowand	CF	32	20%	(0.03)	$ 13,600,000	(0.0025)
Mark DeRosa	LF	35	10%	(0.35)	$ 6,000,000	(0.0577)
Ryan Rohlinger	PH	26	5%	(0.50)	$ 400,000	(1.2500)

Chart 3: Pitching Efficiency

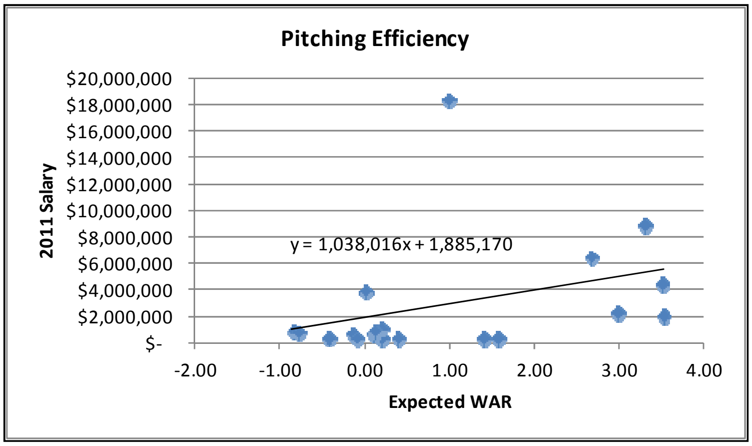

Table 3: Pitching Efficiency

Player	Position	Age	Expected % Season	Expected WAR	Expected Salary + Bonus	Expected Performance Efficiency (per $1m)
Santiago Casilla	RP	30	90%	1.57	$ 420,000	3.733
Sergio Romo	RP	27	90%	1.39	$ 416,500	3.346
Jonathan Sanchez	SP	27	100%	3.52	$ 2,100,000	1.677
Mad. Bumgarner	SP	20	75%	2.97	$ 2,400,000	1.239
Dan Runzler	RP	25	50%	0.37	$ 400,500	0.931
Matt Cain	SP	25	100%	3.50	$ 4,583,333	0.763
Denny Bautista	RP	27	50%	0.18	$ 395,000	0.458
Brian Wilson	RP	28	100%	2.67	$ 6,500,000	0.411
Tim Lincecum	SP	26	100%	3.30	$ 9,000,000	0.367
Ramon Ramirez	RP	28	50%	0.19	$ 1,155,000	0.163
Javier Lopez	RP	32	10%	0.09	$ 775,000	0.113
Chris Ray	RP	28	25%	0.10	$ 975,000	0.107
Barry Zito	SP	32	75%	0.98	$ 18,500,000	0.053
Jeremy Affeldt	RP	31	100%	0.00	$ 4,000,000	-
Guillermo Mota	RP	36	50%	-0.17	$ 750,000	(0.222)
Joe Martinez	RP	27	20%	-0.11	$ 400,000	(0.273)
Todd Wellemeyer	RP	31	50%	-0.86	$ 1,000,000	(0.859)
Brandon Medders	RP	30	50%	-0.80	$ 820,000	(0.976)
Waldis Joaquin	RP	23	10%	-0.43	$ 400,500	(1.070)