Optimizing the First XI:

Salary Allocation Strategies for Major League Soccer

Michael Fotopoulos, Andrew Opatkiewicz
Chicago., IL, USA 60614
Email: fotopnd@gmail.com, andrew.opatkiewicz@prozonesports.com

Abstract

Major League Soccer (MLS) clubs are subject to strict salary and player restrictions limiting roster construction. While a salary cap fosters league parity, it introduces an economic dimension to teambuilding, and gives clubs with more effective resource allocation advantages over their competitors. Traditionally, the fluid nature of soccer, primarily its lack of discrete events, has limited attempts to quantify player performance. We hypothesize that conventional player valuation methods are inefficient means to allocate team salary, and have compiled detailed performance data and salary figures for MLS players for the 2008-2011 seasons. Using linear weighting to quantify player performance beyond traditional measures, we propose disconnects between player performance and player salary. By incorporating more quantitative approaches to player evaluation, we assert that clubs can improve production, optimize limited salary resources, and better inform personnel decisions.

Introduction

Major League Soccer (MLS) has a salary structure unique to the football world. With the league constructed as a single entity, clubs receive a share of a common pool of salary cap dollars to spend on their rosters each season. Though there are various other roster restrictions and exceptions, there is a core issue of salary allocation for any given club's roster decisions. The ability to properly price individual talent with regards to the available player pool should give a club an advantage over its peers. While this is easier said than done, the presence of the salary cap allows us to pivot this question from a qualitative perspective to a quantitative perspective. Whereas performance analysts struggle to search for a unified model for evaluating players, this paper proposes shifting this approach to an economic one. Instead of focusing on which means of production are most useful in projecting team wins, we will be focusing on a player's individual output relative to his peers, combining this with other decision factors like salary, nationality, and academy/international/college experience. The aim here is to move towards a more integrated approach of player evaluation for MLS. Accepting as given that varying styles of play and team composition can influence a player's measurable production, we attempt to focus on game events that isolate a player's on-the-field production. For example, while Player A may have more shot attempts than Player B due to team effects like role, position, or quality of service, we can still measure the players' contribution to his team's overall production through these events. Focusing on effectiveness both on the field and on the club's salary ledger, we look for strategies to maximize the utility of each salary cap dollar spent.

Production vs. Performance

In his 2007 paper [1], Bill Gerrard approaches the question of applying sabermetric principles as seen in Major League Baseball to complex invasion sports such as soccer, rugby, American football, basketball, and ice hockey. The lack of discrete events in these sports combined with the influence of teammates and off-the-ball actions skew measures of individual performance. The current struggle in soccer performance analysis is attempting to isolate player actions and abilities that directly correlate to on-field success much in the way that baseball's on-base percentage, slugging percentage, and the combined OPS+ metrics provide better insight into offensive performance. Given the constraints of the Major League Soccer format, we move the question from projecting player performance towards identifying player production in three phases of the game (offense, defense, and possession) and rating past production against the rest of the available MLS player pool. In the absence of like-for-like events to compare player performance, we look to analyze a player's individual output for a given season and compare the overall percentile rank of that output against all player-seasons for the 2008-2011 data sets.

Borrowing heavily from Branch Rickey's Isolated Power (ISO) [2] and John Hollinger's PER rating methodologies [3], we look to distill a player's ability to convert offensive events into goals into a single measure, Offensive Production (OFF). Through 4 seasons of MLS game data supplied by Prozone Sports Ltd., we observed the frequencies of offensive events and the observed ratios between events. For example, shots on target were 5 times as frequent as goals, so the individual weightings of these events reflect the 1:5 ratio. Breaking this down on an individual level, each player is given a raw score for total offensive production for a given season. These raw scores are then ranked as a percentile among all of the player-seasons for 2008-2011, with the final OFF score for a player season recorded as his percentile rank.

However, as a player's value is not limited to his offensive abilities, we looked to also measure events that would limit opponents' offensive production. Again, utilizing Prozone to isolate player production in the defensive and possession phases of the game, we look to identify match events that contribute to limiting the opposition's ability to turn possessions into scoring opportunities. In the defensive phase, this would be primarily possessions won via tackles or interceptions. In the possession phase (run-of-play), primarily successful passes and headers that sustain possession, contributing to either offensive or defensive play. Following a similar methodology as the OFF score, defensive (DEF) and possession (POSS) scores were then generated to rank player seasons in each phase of the game. These values add to the overall concept that every player contributes to each phase of play over the course of a match.

A full breakdown of these observed event frequencies and the corresponding model weightings can be seen in the Appendix. For more internal consistency, some adjustments were made to the model to account for infrequent events skewing a player's raw score. For example, the frequency of possessions gained to total possessions was observed roughly 2.8 to 1 over a much larger sample size than successful headers. The 2.8 to 1 ratio was used for the less frequent event as both events led to a similar result, the active taking of possession from an opponent.

With each phase of the game scored as a percentile rank, we can derive a summary metric that considers the total output of a player in a given season. Gianluca Vialli provides a simple framework for evaluating players geometrically, taking each phase of the game and plotting it on an x,y plane [4].

While he takes a more qualitative approach to individual metrics, the overall concept is for a player's output to maximize the geometric area of his individual rankings. In our case, the 3 rankings would create a triangle: AREA = .5 * (ATT) * (DEF + POSS). Total AREA scores are then ranked once again as a percentile to view individual player seasons as part of the overall data set. Using this model to quantify player production across all players, we were able to create a general baseline for player valuation. While we do not propose AREA to be a definitive metric for player valuation, a production baseline provides us a frame of reference for analyzing current distribution of salary and playing time/minutes across the player pool and opens the discussion for identifying potential inefficiencies in the player market.

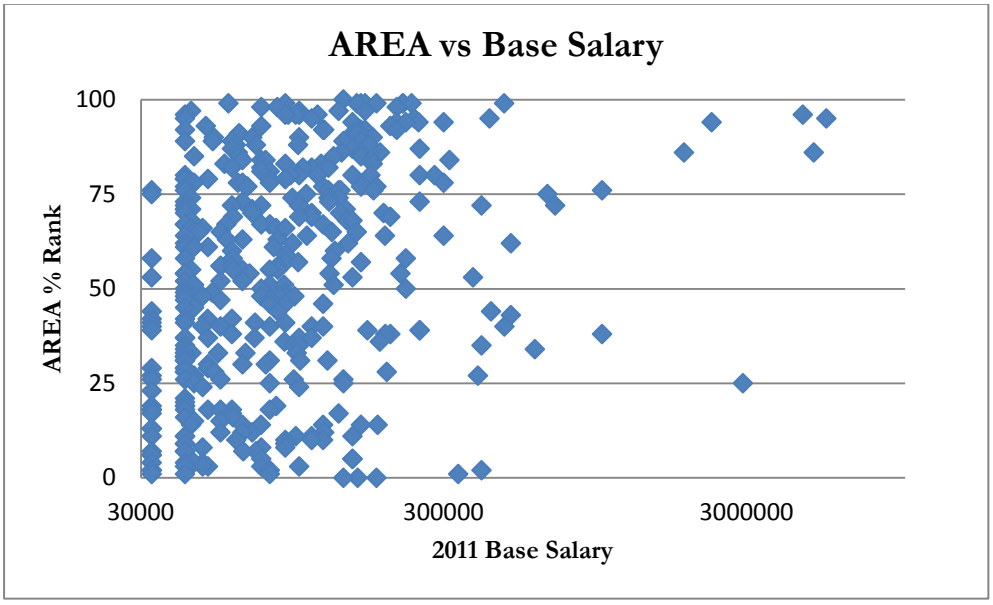


Figure 1 – AREA rankings of player seasons for MLS 2011 Regular Season

Figure 1 illustrates potential inefficiencies in the MLS player market when production is taken into consideration. It should be noted that there are groupings at the left side of the graph are primarily due to the standard rookie contract levels for younger players (under-25) and newly drafted MLS players, $32,600 and $42,000 respectively [5]. With this in mind, there still appears to be a weakness in the relationship between overall player output and base salary. A clearer correlation would remove the clustering seen in the upper-left portion of the graph. The graph indicates an opportunity for cost savings as clubs able to develop methods to exploit any possible mispricing. By taking a closer look at player biographies, we looked to identify trends in the data that could help clubs make better informed decisions for new player signings and roster transactions.

For further research into the 2011 MLS player pool, we recorded available biographical information, and divided the player population into three groups: Academy (IMG Academy or MLS Academy system), College (NCAA), or International (non-US national). While there is some overlap between the 3 cohorts, the intent was to look at each as an individual MLS roster designation. Players primarily enter the league through the academy system, college draft, or as an international signing,

with specific rules and restrictions on each designation. Incorporating AREA scores as a reference point, we were able to take a different perspective on the player pool. Both the cohort information and salary information [6] are given an extra dimension when AREA scores were applied.

Table 1 – Observed Player Cohorts, AREA ratings, and Salary Information for MLS 2011 Season

Player Cohort	Total	+50 AREA score	+80 AREA score
Academy	48	23 (47.92%)	7 (14.58%)
NCAA	218	133 (61.01%)	56 (25.69%)
International	184	97 (52.72%)	39 (21.20%)
Player Cohort	Avg AREA*	Median AREA*	AREA St. Dev*
Academy	47.676	39	29.468
NCAA	53.570	58	31.18
International	48.695	47	30.007
2011 MLS	50.38	52	30.342
Player Cohort	Avg. Base Salary	DP adjusted** Avg.	Median Base
Academy	134,864	90,991	60,188
NCAA	92,846	91,722	70,000
International	238,680	122,754	90,000
Designated Players**	335,000		
2011 MLS League Median Base Salary		160,000	
2011 MLS Avg. Salary		340,755	
DP adjusted Avg.**		104,743	
*minimum 500 minutes played			
** Maximum salary charge for Designated Players is $335,000.			
Adjusted averages reflect salary cap charges			

Conclusions

The AREA model used for the purposes of this investigation is simply one way of interpreting the output of the current MLS player pool. It attempts to provide a comprehensive view of the individual events that contribute to a player's on the field performance, however it does not intend to suggest that maximizing output at the team level will directly impact a team's won-loss record. The events that we were able to capture and our methods of valuing individual events should be

considered flexible for purposes of future analysis. Current data capture techniques are still being developed, and as they continue to evolve, more granular data will become available for performance analysis. Considering the model's application as a basic analytical framework, AREA provides a reference point for evaluating the current player pool and an additional perspective on salary costs and player biographies. MLS roster restrictions place constraints on clubs with respect to both of these issues. Quantifying player output enables clubs to review the allocation of both their cap dollars and international roster slots.

From the data collected, there are a number of observations to be made about the current MLS player pool. Given the similar AREA scores and standard deviations in both the NCAA and International player groups, it would appear that both cohorts provide clubs with similar opportunities to locate productive player talent. In this case, there would be no clear reason to select an international player over an NCAA player simply for his foreign experience. All things being equal, the two player pools provide similar levels of risk and reward, with the international talent coming with higher prices and limited roster space. There are of course exceptions to this concept, clubs with extensive knowledge of specific foreign players will have an advantage over their opponents, but the observed production of international signings overall does not seem to warrant their inflated salary costs.

Identifying college talent should be a top priority for every MLS club. These players have shown to be more productive on average than more expensive international talent. Also, clubs are restricted to eight roster spots for international players, placing further emphasis on using these slots wisely. Both college and academy players are likely to avoid this restriction; only 46 of 450 players were foreign college or academy prospects. Furthermore, top college players receiving Generation Adidas (GA) status for their rookie contracts can provide a double benefit. The GA designation identifies them as the top college talent in a given draft, and their rookie contracts are not counted against a club's salary cap until the player "graduates," shedding the prospect label after a set level of playing time. This effectively gives clubs a "free roll" at top college talent. Continued development of club academy programs will act to supplement college scouting until they are able to fully support first team activities. Until then, the ability to identify MLS-ready talent can pay significant, immediate gains to a club, leveraging the lower cost of entry-level and rookie contracts.

Reducing salary risk while maintaining, or improving upon, player production should be the primary goal of constructing an MLS roster. With more productive minutes purchased with lower rookie and domestic salaries, a club can have more cap and roster resources at its disposal. Considering that cap space is a tradable asset in MLS, clubs can include allocated cap dollars in their transactions, clubs able to conserve salary have more assets with which to deal. Also, reducing the wage bill allows a club to take more risks elsewhere in the roster, a low cost nucleus of players freeing up resources for splashier signings.

Considering higher-priced Designated Players (DPs), it should be noted that these players do not dominate the top percentiles of the observed player-seasons. There are certainly going to be ways these players add value that aren't captured in the current model, Thierry Henry's off the ball movement or David Beckham's field awareness, however the data that we have been able to capture creates an argument that their production can be purchased at lower prices on the domestic player market. However, a player's name recognition value should not be overlooked when considering the

salary cap cost of DPs. As long as these signings can be productive in the upper percentiles of the player pool, their marketability mitigates their increased cap costs.

MLS is a league based on parity through constraint. The basic principles of roster construction are set to prevent any team from spending their way to a title, and to maintain an even, competitive landscape for all 19 clubs. This model has been extremely successful in the National Football League and has contributed to the overall stability of MLS through its first 16 seasons. However, just as elite NFL teams have been able to maintain their success through shrewd management of the salary cap, MLS teams must be able to similarly inform their personnel decisions. Current performance analysis practices are just the beginning of this conversation. Our analysis suggests that there is still room for improvement for clubs looking to make every salary cap dollar as productive as possible. Clubs should be able to exploit the potential inefficiencies we observe in the MLS player market by developing techniques to more effectively price available talent. Our AREA model provides one perspective on the production of the current MLS player pool, and provides a framework for quantitative evaluation. Further refinement of this approach should be considered for any club looking to optimize its salary and roster resources. Though MLS aims to even the playing field through its restrictions, we suggest clubs can use these restrictions to create advantages over their opposition. By minimizing the effects of these league constraints, teams are able to free up more resources with which to compete, improve, and succeed.

References

1. Gerrard, Bill "Is the Moneyball Approach Transferable to Complex Invasion Team Sports?" *International Journal of Sport Finance* 2.1 (2007)**:** 214-230. Print.
2. Rickey, Branch, "Goodbye to Some Old Baseball Ideas," *Life*, 2 Aug 1954: 81. Print.
3. Hollinger, John. "Calculating PER." *Basketball-Reference.com*. Web.
4. Vialli, Gianluca, and Gabriele Marcotti. *The Italian Job: A Journey to the Heart of Two Great Footballing Cultures*. London: Bantam, 2006. Print.
5. Major League Soccer. "2011 MLS Roster Rules." *MLSsoccer.com*. Web.
6. MLS Players Union. "Player Salary Information." *MLSPlayers.Org*. Web.

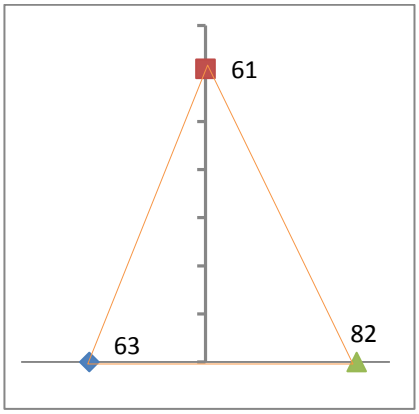
Appendix

Observed Event Frequencies in 2008-2011 MLS Seasons with AREA Model Weightings

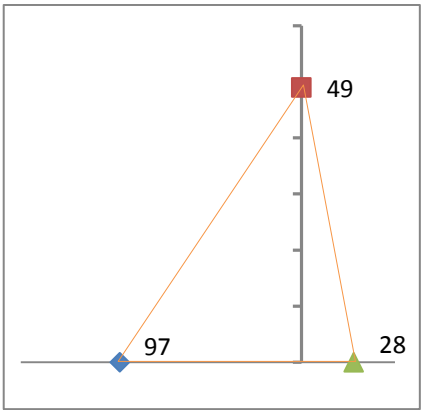
ATT Components	Total	Ratio to Goals	Weighting
Goals	2398	1.000	5.000
ShotsOnTarget	11956	0.201	1.003
1st_Assists	2073	1.157	5.784
Successful Penalty Area Entries	18947	0.127	0.633
Offsides	4830	0.496	-2.482

		Total	Ratio to Possession		Weighting
	Possessions	894630	1.000	n/a	
DEF	Possession Gained (tackles and interceptions)	312343	2.864		2
	Fouls in Defensive 3rd	5023	178.107		-2
	Cross Defended	19591	45.665		2.5
POSS	Successful Short Pass (0-10m)	121359	7.372		1
	Succ. Medium Pass (10-25m)	273977	3.265		1
	Succ. Long Pass (+25m)	79387	11.269		2
	Possession Lost	396750	2.255		-2
	2nd Assist	1600	559.144		2.5
	Successful Header	53582	16.696		2.8

Example of AREA calculation



Stephen McCarthy 2011 (61 ATT, 63 DEF, 82 POSS)



AJ Soares 2011 (49/97/28)

-AREA ratings (.5*ATT*(DEF+POSS)) : McCarthy 2011 -> 4422.5, Soares 2011 -> 3062.5
-AREA scores ranked across all player seasons: McCarthy 84th percentile, Soares 67th

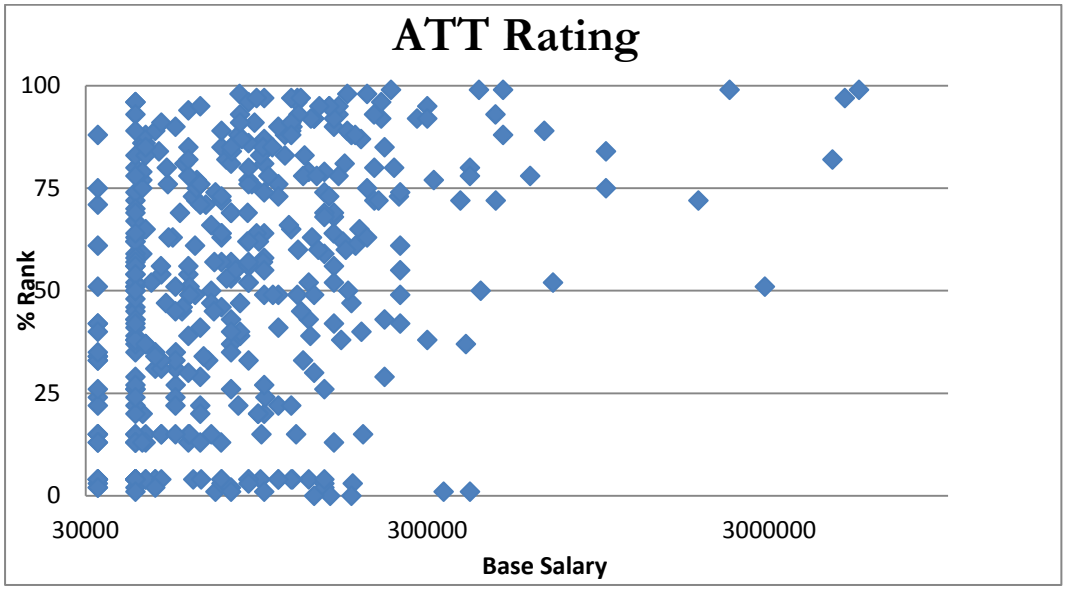


Figure 2 - ATT rankings of player seasons for MLS 2011 Regular Season

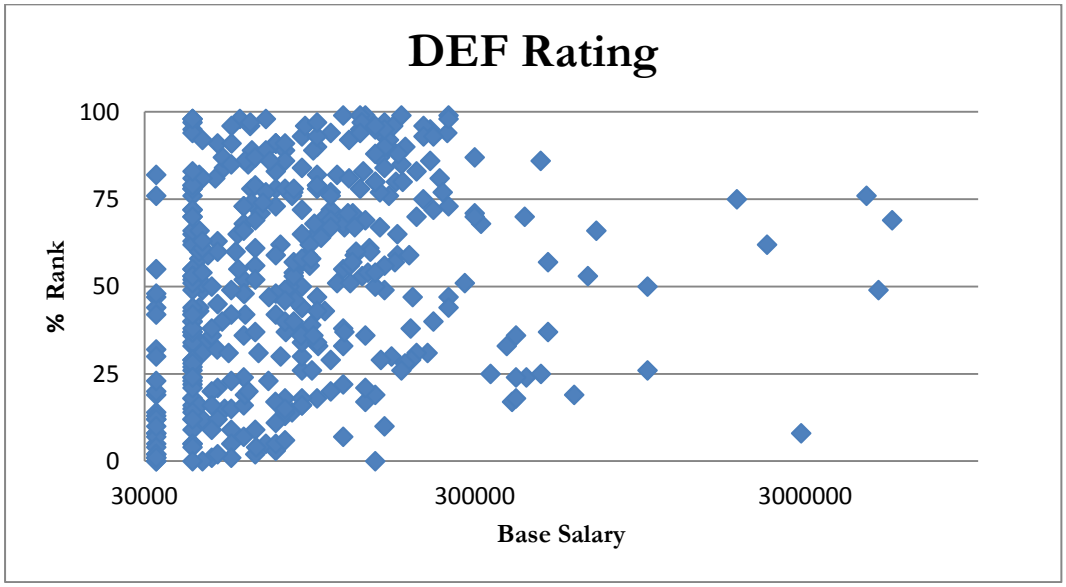


Figure 3- DEF rankings of player seasons for MLS 2011 Regular Season

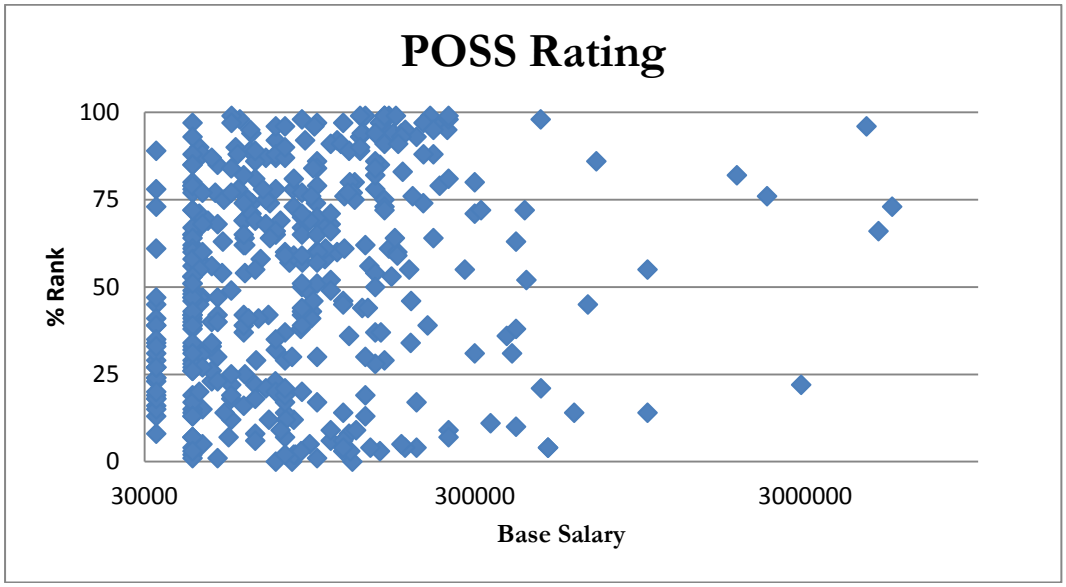


Figure 4 - POSS rankings of player seasons for MLS 2011 Regular Season