

Assessing Value of the Draft Positions in Major League Soccer's SuperDraft

Tim B. Swartz, Adriano Arce and Mohan Parameswaran *

Article Type: Research Article

Article Category: Sports Studies

Running Head: Draft Position Value in MLS

*Tim Swartz (corresponding author: tim@stat.sfu.ca) is Professor in the Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Swartz has been partially supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). Adriano Arce and Mohan Parameswaran are undergraduate students at Simon Fraser University (SFU) who have been partially supported by NSERC's USRA program and SFU's VPR USRA program respectively.

Assessing Value of the Draft Positions in Major League Soccer's SuperDraft

Abstract

This paper assigns various performance measures to players who have been drafted in the SuperDraft of Major League Soccer. These measures are then used to assess the value of draft position. As a by-product of the analysis, we provide an estimate of the trajectory of player performance in a soccer career. Both the valuation of draft position and the career trajectory analysis may be important tools for general managers in Major League Soccer.

Keywords: draft order value, Major League Soccer, sports analytics, SuperDraft.

1 INTRODUCTION

By all measures, soccer (or football as it is more commonly known) is the most popular sport in the world (The Economist 2011). In the United States and Canada, the highest level of soccer is played in Major League Soccer (MLS). As of 2012, MLS consisted of 19 clubs divided into the Eastern Conference (10 teams) and the Western Conference (9 teams). Unlike the traditional European fall-to-spring schedule, the MLS season begins in March and ends in November. The MLS team with the best regular season record is the recipient of the *Supporters Shield* and the winner of the MLS playoffs receives the *MLS Cup*.

There are various ways that MLS teams acquire players for their rosters. One such mechanism is the annual *SuperDraft* whereby players are placed on the draft-eligible list via nomination from MLS clubs. Most of these players are US college players. Players are then drafted by MLS teams in the inverse order of team performance from the previous season. The ordering is an attempt to improve competitive balance within the league. The number of players drafted has varied since the inception of the SuperDraft in 2000. In the most recent draft of 2012, 38 players were drafted (two rounds involving the 19 MLS clubs).

The primary problem considered in this paper is the valuation of draft order in the SuperDraft. This is a fundamental problem for general managers in MLS. Consider the following hypothetical scenarios where the knowledge of draft value would be useful:

- Your team has a glaring weakness at fullback and you have the 20th pick in the draft. What are the chances that the draftee can satisfy your needs at fullback? Would it be a better strategy to fill this position via a trade or via a designated player?
- Your team is facing salary cap constraints and you have the 10th pick in the draft. What is the salary expectation for the player?
- Your team has the draft rights for both the 10th and 11th picks. Would it be in your interest to trade both of these picks for the first draft pick?

The valuation of draft order was first considered in the context of the annual draft of the National Football League (NFL). Armed with a draft value chart, Jimmy Johnson (coach of the Dallas Cowboys) made informed trades on draft day in 1991, acquiring a whopping 19 players for the Dallas Cowboys. These informed trades eventually turned the team's fortunes around as Dallas won three Superbowls within the next five years. Moskowitz and Wertheim (2011) provide an entertaining account of the use of draft value charts in the NFL.

The early chart was constructed according to the value of draft picks based on actual trades that took place. Massey and Thaler (2010) investigated the chart, and using data based on future salary contracts, determined that the early chart overvalued high draft picks. For example, although the first draft pick is valuable, the early chart suggested that the first pick is more valuable than is actually the case. Shuckers (2011) corroborated the overvaluation associated with the early draft chart via alternative measures of player performance. Shuckers (2011) subsequently proposed an alternative draft value chart. In other sports with drafts, Berry (2001) has looked at the success of first round picks. However, it does not appear that there has been any published investigation of the value of draft order in the MLS SuperDraft.

One of the complicating factors in assessing the value of draft order in the SuperDraft is that many of the players who have been drafted have not yet reached their peak level of play. Although some players when drafted in their early 20's are unable to make an immediate contribution, through training and experience, they eventually become serviceable MLS players. We therefore want to be able to assess the future value of draft picks. In section 2, we address this problem by estimating career trajectories of soccer players. Common sense dictates that players generally improve early in their career, then peak, and finally experience a decline in performance. The age at which they peak, and the extent of their improvement/deterioration around the peak is the focus of section 2. Careful data collection is required to assess player productivity. Although the estimation of career trajectories is a necessary step in assessing draft order value, the problem itself is one of considerable interest

to general managers. For example, what length and value of a contract should be offered to a player who is two years past his peak?

In section 3, we assess the value of draft position. The primary difficulty in the exercise is the determination and collection of performance measures. Clearly, a common metric such as “goals scored” is not relevant to all positional players. For example, defenders are not expected to score goals. The end product of our analysis are graphs of value plotted against draft position. Various graphs are provided based on various performance measures. With such graphs, one can assess for example, the relative value of the 20th draft pick to the first draft pick. We conclude with a short discussion in section 4.

2 CAREER TRAJECTORIES IN SOCCER

Assessing performance in soccer is not a straightforward task. For example, although *goals scored* is a popular and important statistic, it is not relevant to all positional players, especially defenders. Also, imagine a player who changes teams and experiences a surge in productivity. This may have more to do with circumstances surrounding his new team than an improvement in his personal play. In this section we make a number of subjective and hopefully reasonable decisions with respect to assessing player performance. Yearly performance data coupled with age data will help us to evaluate career trajectories in soccer.

We began by considering players who have played on top flight clubs for a period of at least five consecutive seasons. We have chosen 12 teams from Europe which have been traditional footballing powers. The rationale is that these clubs are of the highest quality and consistency. Therefore an observed change in performance when competing for these clubs is assumed to convey a change due to the player. We have restricted our analysis to the “modern” era of soccer beginning with the 1992/1993 season when the English Premier League was formed. We have excluded goalkeepers from the analysis as it is well known that keepers can remain competitive at more advanced ages. For example, Edwin van der Sar

retired as keeper from Manchester United following the 2010/11 season at 40 years of age. Although players from the 12 chosen clubs are of exceptional quality, we assume that the average career trajectory for players at these clubs is not unusual. For example, this implies that the average peak age for players at the elite clubs should be the same as the average peak age for all professional soccer players.

From the websites www.footballdatabase.eu and www.bdfutbol.com/en/e/e.html we identified 232 players who met our criteria. These players played a total of 1791 seasons. We were also able to collect seasonal performance data on each of the players in terms of *minutes played*. Although *minutes played* does not capture all of the elements of performance, it is clearly a sensible measure of worth. The data are summarized in Table 1.

Team	League	No. Players
Arsenal	English Premier League	25
Chelsea	English Premier League	11
Liverpool	English Premier League	23
Manchester United	English Premier League	25
Bayern Munich	German Bundesliga	26
AC Milan	Italian Serie A	20
AS Roma	Italian Serie A	18
Internazionale	Italian Serie A	14
Juventus	Italian Serie A	18
Barcelona	Spanish La Liga	20
Real Madrid	Spanish La Liga	16
Valencia	Spanish La Liga	16
Overall		232

Table 1: Summary data for players who have played at least five consecutive years with a top flight club sometime during the 1992/93 through 2011/12 seasons.

Recall that our objective in this section is the evaluation of career trajectories. Therefore, we want to assess individual player performance at different ages where player performance is measured against himself. Accordingly, let x_{ij} denote the minutes played for the i th player

at age j in the season where the player was j years old on January 1. The beginning of January is roughly halfway through the traditional fall-spring soccer season. Corresponding to x_{ij} , let w_{ij} be the number of available minutes during the season where w_{ij} is the product of 90 minutes and the number of games that his team played in the season. We then define $z_{ij} = x_{ij}/w_{ij}$ as a performance statistic for the i th player in the given season as it represents his proportion of available minutes on the field. To provide a personal measure of performance for the i th player at age j , we define

$$pp_{ij} = \left(\frac{z_{ij}}{\max_j \{z_{ij}\}} \right) 100\% \quad (1)$$

where the maximum is taken over all of his active seasons at the club. Therefore the personal performance measure pp_{ij} for the i th player has a maximum value of 100% in at least one of his seasons. Note that we only considered matches played in the domestic season. Our rationale for this choice is that other matches (e.g. friendlies, FA Cup, Europa League, Champions League, etc) may not have a consistent level of competition. Also, it is well known that players are often given rest when matches are scheduled in frequent succession.

In Figure 1, we provide a scatterplot of the personal performance measure pp_{ij} versus age j based on *minutes played* for all players over all seasons. The plot consists of 1791 points. Although there is considerable variability in the scatterplot, the lowess function describes the overall trend. The shape of the lowess plot corresponds to our intuition where performance improves early in a career, then peaks, and concludes with a period of decline. We observe that typical player performance peaks at roughly 24-27 years of age. Also, the dropoff is minimal for roughly two years surrounding the peak period. The improvement early in a career is more rapid than the decline at the end of a career. We note that although the peak period of 24-27 years appears to be in rough agreement with common opinion (www.forum.ea.com/uk/posts/list/802432.page), we know of no quantitative study that has provided such an interval. A miscalculation by even a couple of years can have dire consequences for general managers when deciding upon long term contracts. The parameters

of the lowess function were set according to $\text{span}=0.5$ and $\text{degree}=2$ which is consistent with the analyses of Shuckers (2011). The choice of the span parameter appears reasonable as we do not want large fractions of the data to fit local segments of the trajectory curve. For example, the trajectory at young ages does not likely have much to do with longevity. The degree parameter should also take a value exceeding 1 since we do not expect linear segments in the trajectory curve. Setting $\text{degree}=2$ provides more curvature.

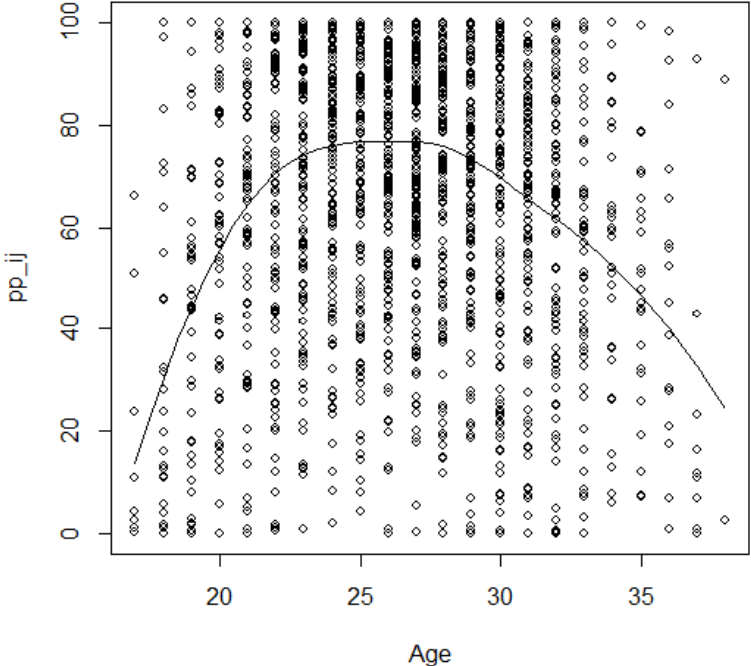


Figure 1: Scatterplot of the personal performance measure pp_{ij} based on *minutes played* versus age. A lowess fit is included to help assess the overall trend.

Recall that each player has at least one plotted value of 100%. Accordingly, there is an inherent assumption that data are collected over a period that includes each player’s peak years of play. The assumption is valid for most players since it is unlikely that a player had five consecutive years of form that were all below peak performance, yet he managed to play

on one of these top flight teams. However, to check the robustness of the assumption, we have repeated the analysis based on a restricted subset of 129 players who played at least seven consecutive seasons with one of the 12 top flight clubs. These players played a total of 1229 seasons. A greater period of seasons improves the chance of capturing a player's peak year. We found no meaningful difference in the resultant plot when compared to Figure 1. Related to the assumption, we encountered 9 players (e.g. Zinedine Zidane) who played at least five years on each of two teams from our list. In these cases, we only retained the years at the club where the player was 26 years of age. In the case of Zidane, we used his Juventus years instead of his years at Real Madrid.

How do we interpret the vertical scale in Figure 1? From 24-27 years of age, an average player operates at roughly 75% of top performance. Similarly, an average 34 year old operates at roughly 50% of top performance. Therefore, we might view the 35 year old as being able to contribute $50/75 \rightarrow$ two thirds of what he was able to contribute at his peak. Here, the measure of contribution can be interpreted in minutes played.

3 ASSESSING VALUE OF DRAFT POSITION

In this section we investigate the relationship between player value and draft position in the MLS SuperDraft. To facilitate the investigation, www.wikipedia.org/wiki/MLS_SuperDraft provides the entire history of the MLS SuperDraft going back to the inaugural draft on February 6, 2000. The list consists of a total of 745 players chosen in the 13 drafts from 2000 through 2012. The number of players drafted per year ranges from 72 (2001) to 38 (2012).

Whereas obtaining the draft position of each player is uncontroversial and routine, the determination of player value is far from straightforward. As in section 2, we first consider value measures based on minutes played. Another indicator of player value which we consider is yearly salary. Although players can be underpaid or overpaid in a given contract, subsequent player contracts tend to adjust to reality. A difficulty with using

salary as a proxy for performance is that there are time lags between observed performance and contract. Fortunately, for the MLS, there are good sources of data. For minutes played, we made use of <http://socceroutsider.com/> and player's personal Wikipedia sites. For salary information, data are available for the six seasons 2007 through 2012 at www.mlplayers.org/salary_info.html. We had to dig deeper for earlier seasons, referring to www.washingtonpost.com/wp-srv/sports/mls/longterm/2006/mls.salaries.html for the 2006 season, www.bigapplesoccer.com/article.php?article_id=3103 for the 2005 season and <http://sportsillustrated.cnn.com/2004/soccer/01/06/mls.salaries.sa/> for the 2004 season.

Before the various value metrics are introduced, we indicate some general problems that need to be addressed with respect to player valuation:

- A spectacular season is valuable to a club. There is also value in a longstanding career. How do we balance short term performance with longevity?
- How do we assess a player who was drafted in the SuperDraft but went on to play in some other league? For example, Clint Dempsey was drafted 8th in the 2004 SuperDraft by the New England Revolution. However, since 2006/07, Dempsey has played in the English Premier League for both Fulham and Tottenham. Clearly, Dempsey is a valuable player although he has only limited MLS data.
- In our current list of drafted players, some players are still young and have not yet reached their full potential. Should we assess their value on some combination of current performance and future performance? The career trajectory plot Figure 1 may be helpful in this regard.
- With the growth of the MLS, there has been an escalation in salaries over time. How do we compare a salary from the early years to recent salaries?

To facilitate a comparison of various metrics, we define each metric on a scale of 0 to 100.

We begin by considering minutes played in regular season MLS matches. The restriction to regular season matches levels the comparison amongst players since each team has the same number of games of comparable importance. For each player, we calculated his percentage of total minutes played relative to available minutes in a season. This was done for each year over all of his MLS seasons. We then defined the player performance metric y_1 for a given player as his maximum percentage over his MLS career. We also defined the player performance metric y_2 for a given player as his average percentage over his three best MLS seasons. Therefore, the measure y_2 favours career longevity over y_1 .

In the case of players who have gone on to play in “superior” leagues, we arbitrarily assign a percentile rank of 90% in seasons where they played in superior leagues. We define a superior league as any of the four famous leagues listed in Table 1. Players who have made it to one of these four top flight leagues have typically developed in the MLS before making their jump to the big time. Although there are other leagues that many football fans would agree are better than the MLS (e.g. Liga Portuguesa (Portugal), Ligo Do Brasil (Brasil), Ligue One (France), Primera Liga (Argentina), Eredivisie (Holland)), very few MLS drafted players have gone on to play in these professional leagues. When an MLS draftee plays in leagues other than the MLS or one of the four top flight leagues, we assign a percentile rank of 0% for those seasons. A 0% score is also assigned to players who discontinue playing. A rationale for the 0% performance score is that it was a mistake to draft such a player as they did not contribute to the MLS team that drafted them. In our list of MLS drafted players, only 12 played in superior leagues. Of the 339 that played in other leagues, 277 played in the NASL (North American Soccer League) or the USL (United Soccer Leagues) which can truly be viewed as lower quality leagues compared to the MLS. When a player belonged to multiple leagues in a year, we used the league where he played most of his games.

To account for young players who have not yet reached their top level of performance (i.e. less than 24 years of age), we imputed values for their unobserved seasons. For example, suppose that we have a 21 year old player in the MLS who has just completed a season. We

take his minutes played as a 21 year old and multiply by $l(22)/l(21)$ to obtain his predicted minutes as a 22 year old where $l(x)$ is the value of the lowess function at age x in Figure 1. We do not allow predicted percentages to exceed 100%. The determination of player ages was not as straightforward as we had hoped; there were at least six additional websites that we accessed to collect birthdates.

An obvious difficulty with the minutes played variables y_1 and y_2 is that they do not account for team strength. For example, it is easier to play extended minutes on a poorer performing club than at a stronger club. Another difficulty with these measures is that there are a number of players who represent their club in nearly all of the regular season matches. Therefore minutes played does not adequately distinguish these players in terms of performance.

Our preferred performance metrics are based on salary data. Varying team strength is less of an issue when dealing with salary data since league-wide salary caps exist. Salary caps help to impose a realism on salaries so that players are paid what they are worth. An exception are the salaries paid to designated players who are not MLS draftees. For each player, we calculated his percentile rank for a given year based on his salary relative to all MLS players in that year. Thus the comparison sensibly involves the performance of draftees against the population of MLS players. This was done for each year over all of his MLS seasons. We then defined the player performance metric y_3 for a given player as his maximum percentile rank over his MLS career. We also defined the player performance metric y_4 for a given player as his average percentile rank over his three best MLS seasons. We used the same rules as above in handling players who have gone on to play at superior clubs and for young players who have not yet reached their full potential. In the case of multi-year contracts, the amount that a player received in a year is used as his yearly salary.

In Figure 2, we provide a scatterplot of the preferred y_4 metric versus draft position. To assess the overall trend, a lowess plot is superimposed. We observe the anticipated pattern that early draft picks have more value. The plot decreases rapidly during the early picks

and then levels off. It is interesting that there is little additional value beyond draft position 25. What this suggests is that whereas general managers have good intuition of value early in the draft, late draft picks are more or less a crapshoot. It also suggests for example that managers should value a 50th draft pick as about equal to a 25th draft pick. If this sort of information is not well understood, a savvy manager may be able to trade a 25th draft choice for a 50th draft choice plus additional assets. We observe from the variability in the plot that it is still possible to draft a productive player late in the draft but the probability of doing so is much decreased. Moreover, the variability is greater for early draft picks. This suggests that there may be great pressure for teams drafting early as an early draft choice can turn out to be either a star or a bust. With less expected of late draft picks, there may only be upside for a general manager. How might we interpret the vertical scale of the plot? According to the salary metric, the first draft pick provides you with a player who on average ranks in the top 85.6% of players in the MLS. The first draft pick is about 1.7 times as valuable as the 10th draft pick and is about 5.3 times as valuable as the 20th draft pick in terms of salary. Finally, we note that we obtain a very similar plot if y_4 is based on the best four seasons rather than the best three seasons. For ease of reference, the values of the loess curve in Figure 2 are provided in the pick value chart given in Table 2 of the Appendix.

In Figure 3, we provide a comparison of the lowess plots using each of the four proposed metrics. We observe that each plot conveys similar information and this provides assurance that the valuations are meaningful. For example, all four plots indicate that the first draft pick injects a team with a player who on average will rank roughly in the top 80% of MLS players. The level of agreement in the four curves was a bit of a surprise to us as we were aware of the flaws in assessing value based on minutes played. For the most part, we also note that y_1 dominates y_2 and that y_3 dominates y_4 . This suggests that the consideration of a single season (y_1 and y_3) tends to inflate a player's value when compared to their performance over multiple seasons (y_2 and y_4). Finally, we note that the preferred metric y_4 lies amongst the middle of the four lowess curves.

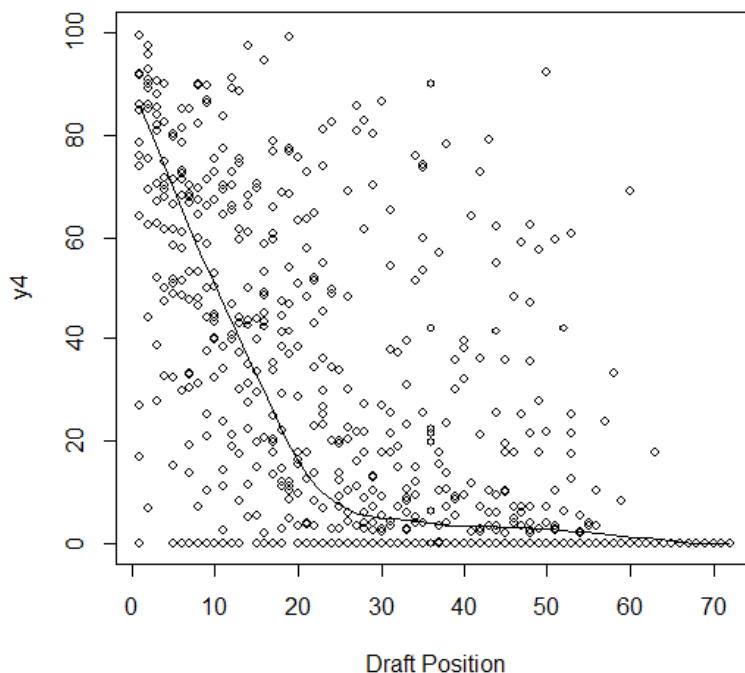


Figure 2: Scatterplot of the value metric y_4 versus draft position. A lowess fit is included to help assess the overall trend.

4 DISCUSSION

The major contribution of the paper is the construction of Figure 1 and Figure 2. In Figure 1, we have estimated the performance trajectory of soccer players (excluding keepers). The plot may be of value to general managers in planning team rosters and offering contracts. In Figure 2, we have estimated the value of draft picks in the MLS SuperDraft. This may also be of value to general managers when planning rosters and assessing trades.

The major difficulty in both the construction of Figure 1 and Figure 2 is the determination of player value. Value is a subjective quantity and involves many factors including player age, the assessment of the importance of longevity, positional characteristics, the confounding of

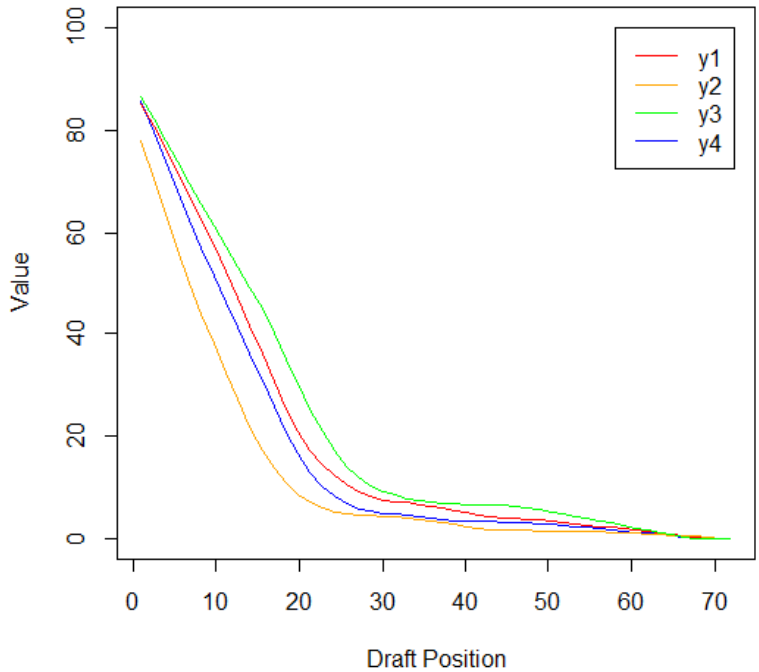


Figure 3: Lowess plots for each of the value metrics.

individual and team characteristics, changes in league salaries and missing data. We have attempted to handle these issues sensibly and we note that various value metrics have lead to similar results (see Figure 3). Data collection and data management also proved to be a substantial exercise in obtaining our results. It may be interesting to extend the work by restricting analyses to various positions (e.g. defenders, midfielders, forwards) although this reduces the sizes of datasets and often introduces uncertainties with respect to the categorization of players.

Finally, a relatively new change may affect the future of the MLS SuperDraft. In 2006, the Home Grown player criteria was established in the MLS to nurture promising young players living in the vicinity of an MLS team. The designation requires commitment by

players to practice/play sufficiently under the club’s development system. Teams may then sign a player to their first professional contract if the player has trained for at least one year in the club’s youth development program and has met the league’s Home Grown player criteria. The important implication is that such players do not participate in the SuperDraft. Therefore, although the Home Grown player program is currently in flux, should it gain widespread popularity, it may eventually dilute the quality of the SuperDraft.

5 APPENDIX

Pick	Value	Pick	Value	Pick	Value	Pick	Value
01	85.6	19	18.8	37	03.6	55	02.0
02	81.5	20	16.1	38	03.5	56	01.9
03	77.5	21	13.5	39	03.5	57	01.7
04	73.5	22	11.3	40	03.4	58	01.6
05	69.5	23	09.6	41	03.3	59	01.4
06	65.6	24	08.3	42	03.3	60	01.2
07	61.7	25	07.3	43	03.3	61	01.1
08	57.9	26	06.3	44	03.2	62	00.9
09	54.2	27	05.7	45	03.1	63	00.7
10	50.5	28	05.4	46	03.1	64	00.6
11	46.8	29	05.1	47	03.1	65	00.4
12	43.2	30	04.8	48	03.0	66	00.2
13	39.7	31	04.7	49	02.9	67	0.00
14	36.1	32	04.6	50	02.7	68	0.00
15	32.7	33	04.3	51	02.6	69	0.00
16	29.2	34	04.1	52	02.5	70	0.00
17	25.4	35	04.0	53	02.3	71	0.00
18	21.9	36	03.8	54	02.2	72	0.00

Table 2: Pick value chart corresponding to the lowest curve in Figure 2.

6 REFERENCES

- Berry, S.M. (2001). “Do you feel a draft in here?”, In the column, “A Statistician Reads the Sports Pages”, *Chance*, 14(2): 53-57.
- Massey, C. and Thaler, R.H. (2010). “The loser’s curse: overconfidence vs. market efficiency in the National Football League draft”, <http://ssrn.com/abstract=697121>.
- Moskowitz, T.J. and Wertheim, L.J. (2011). *Scorecasting: The Hidden Influences Behind how Sports are Played and Games are Won*, Crown Archetype: New York.
- Shuckers, M. (2011). “An alternative to the NFL draft pick value chart based upon player performance”, *Journal of Quantitative Analysis in Sports*, 7(2): Article 10.
- The Economist (2011). Game Theory blog post under “Ranking sports’ popularity: And the silver goes to ...”, www.economist.com/blogs