Great Expectations: An Analysis of Major League Baseball Free Agent Performance

Sean L. Barnes[1]*, Margrét V. Bjarnadóttir[1]

[1] *Department of Decision, Operations & Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742-1815, USA*

Received 30 June 2015; revised 9 December 2015; accepted 15 February 2016
DOI:10.1002/sam.11311
Published online 29 June 2016 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: We explore whether free agents in Major League Baseball meet the expectations set forth by newly signed contracts. The value and duration of these contracts are negotiated between the player (and his agent) and the signing team and are based primarily on the player's performance to date, projected future performance, and potential marketing value to the team. We develop two classes of models to explore this problem using a variety of regression- and tree-based machine learning algorithms. The market model uses player and team data to predict the market value of a player's performance (i.e., average contract salary). The performance model uses the same data to predict wins above replacement as a surrogate for overall player performance. We translate this measure into dollars using position-based conversion factors. Analysis of these models demonstrates that the performance model more consistently predicts and assesses player value with respect to their free agent contracts. Together, these models can be used to target or avoid free agents (or other players) whose performance-based value differs significantly from their market value. Statistical Analysis and Data Mining: The ASA Data Science Journal 9: 295–309, 2016

Keywords: baseball; analytics; machine learning; compensation; free agents

1. INTRODUCTION AND MOTIVATION

Major League Baseball (MLB) General Managers (GMs) and their staffs are charged with identifying and acquiring players who will contribute positively to team success. An open question in MLB (and many other professional sports) is which team building strategy is optimal for maximizing the potential for winning. Does spending more money lead to a higher likelihood of winning? In recent years, large-market franchises such as the New York Yankees, Boston Red Sox, Los Angeles Dodgers, and Los Angeles Angels have had significant success with spending a lot of money on free agent players, whereas smaller-market teams such as the Oakland Athletics, Tampa Bay Rays, Pittsburgh Pirates, and Kansas City Royals have found success with an approach focused on spending money more efficiently. In Fig. 1, we visualize the relationship between team payroll and wins for the 1998–2013 MLB seasons, highlighting the teams that qualified for the playoffs. It is clear from this figure that teams can have successful (i.e., winning) seasons with significantly different team-building philosophies in terms of payroll.

Regardless of the overall team strategy, all teams prefer to sign players who will perform well with respect to the newly agreed upon contract. When attempting to sign a player via free agency, the GM must determine the value of that player in terms of their potential contributions on and off the field and then negotiate a contract with the player's agent. In an objective world, player contracts would be solely based on a player's expected future performance, but there are other factors that can play a significant role, such as a player's marketability or competitive pressure from other teams.

MLB fans are constantly (and subjectively) evaluating a player's on-the-field performance. However, it is uncommon to see an objective evaluation of how well free agents meet the expectations set by their contracts. Researchers have previously analyzed how player performance translates into compensation. Several efforts have focused on evaluating the Moneyball hypothesis, that is, how on-base percentage and slugging percentage translate to player salaries [1–3]. Rockerbie developed a more robust

* *Correspondence to:* Sean L. Barnes (sbarnes@rhsmith.umd.edu)

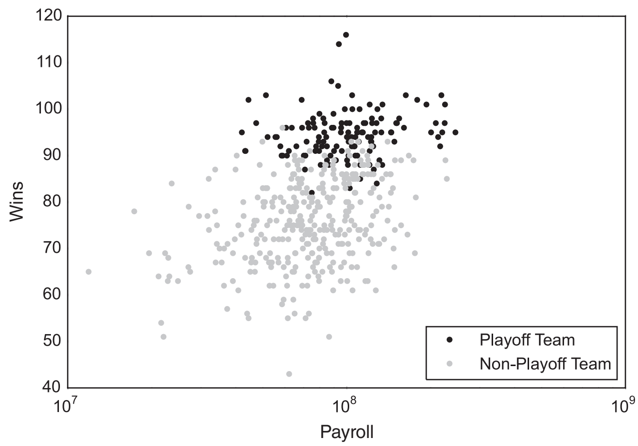

Fig. 1 Comparison of team wins versus payroll from 1998 through 2013.

regression model of free agent salaries based on auction theory [4]. Turvey explored the benefits and risks of long-term contracts for younger players [5] but only evaluated two specific contracts. Despite these contributions, teams still lack methods to project players' future performance accurately and use those projections to make effective decisions about contract durations and values.

In this study, we build models to predict and assess players' market and performance-based value. We specifically evaluate players' market value based on free agent market conditions, that is, the market value of a player given his performance to date. For the performance model, we evaluate a player according to the value of his expected future wins above replacement (WAR), also based on his performance to date. WAR represents the contribution of an individual player to team wins—relative to a replacement-level player—and has become the standard for a composite measure of player performance; therefore, accurate predictions of WAR form a viable basis for value-driven player compensation. A key advantage of using WAR for the performance model—over other advanced metrics (e.g., win probability added)—is that it also incorporates defensive performance.

Next, we compare and contrast the market and performance models. First, we explore how players' performance to date affects these two estimations (i.e., market value and expected future WAR). In other words, which statistics contribute positively toward player value, which statistics contribute negatively toward player value, and which statistics have little-to-no effect? Second, using each model, we assess how often players meet the expectations set by their free agent contracts. Third, we evaluate both approaches and identify the approach that is more reliable for projecting future player value. And lastly, we combine the two approaches and retrospectively identify players whose predicted performance value differed significantly from their market value. GMs that operate under monetary constraints will be interested in players whose expected performance value exceeds that of their market value.

Our study is unique in that it incorporates substantially more data than many previous studies, both in terms of the number of players and the number of statistics included for each player. We consider several analytical methods, each with different characteristics in terms of explanatory and predictive power, and we compare our market and performance valuations to complete contract data instead of simply yearly salaries. Finally, we devise a new method for identifying potentially undervalued and overvalued players.

2. DATA AND DATA MODELING

Player performance data was acquired via Baseball-Reference for both position players and pitchers (starters and relievers) who played between 1998 and 2014 [6]. The data contains information on the position, age, year, team, league, and salary for all players in each included season. We adjusted all player salaries to 2014 U.S. dollars. The batting data contains detailed information on each player, including basic and advanced batting statistics, batting ratios, and career statistics (including cumulative, average, and playoff statistics). Detailed situational batting, base running, and fielding data were not collected. The pitching data similarly contains detailed information, including basic and advanced pitching statistics, pitching ratios, batting against statistics, starting and relief pitching statistics, and career statistics (again including cumulative, average, and playoff statistics). Detailed situational pitching and pitch data were not collected. A large number of baseball statistics are strongly correlated. For example, games played, plate appearances, and at bats are nearly perfectly correlated for batters. Innings pitched and batters faced are similarly interchangeable for pitchers. In order to facilitate comparisons across different models, we removed approximately 30 of these strongly correlated variables for batters and 20 for pitchers. In addition to the player performance data, we included player contract and transaction data from three sources, including ESPN's Player Tracker [7], Cot's Contracts [8], and RETROSHEET [9]. We also attempted to infer missing contract information based on free agent signings reported on Baseball-Reference as this information was sparse and fragmented for the earlier years in our dataset. We note here that we reference each contract in the text using the original terms, but we used adjusted figures for all of the models. Finally, we include team-specific information for each season, including the win–loss record, runs scored and allowed, playoff results, and payroll and attendance as surrogates for market size.

We filtered our batting and pitching datasets to only include players eligible for free agent contracts. These are essentially players whose salaries are decided on the basis of the player's performance to date, the potential for future performance, and the marketability and long-term value of the player to the franchise. We excluded players in the first 2 years of their major league career because their salaries automatically renew for each of their first 3 years. We also excluded players who were eligible for salary arbitration (players with 3–5 years of experience) because they were also not eligible to receive contracts fully representative of their market value. Exceptions to these exclusions include arbitration-eligible players who received multiyear contract extensions. For example, Philadelphia Phillies 2B Chase Utley signed a 7-year/$85M contract after only his third major league season in 2006. We set a minimum of 130 at bats for batters and 50 innings pitched for pitchers in order to filter out players who did not play frequently at the major league level. Players who began their careers in foreign professional leagues were only included when historical MLB data was available. In total, we trained our models on data from 1727 player-seasons for batters and 1256 player-seasons for pitchers. Due to the limited number of players and high number of variables, we elected to use 5-fold cross-validation for model evaluation instead of a fixed partitioning scheme.

3. METHODS

We train and cross-validate the market and performance models on the full filtered batting and pitching datasets and evaluate them based on out-of-sample observations. The dependent variables are the average yearly salary for the new contract (i.e., the full contract value divided by the duration of the contract) for the market model and the next year's WAR for the performance model. Using average new contract salary enables the comparison of players who signed different contracts of different lengths and has the advantage over next year's salary of not being subject to the annual variation common to multiyear contracts. We evaluate the predictive performance of both linear and log-transformed models for average new contract salary.

We build models for batters and pitchers independently. All models predict the average yearly outcome for the duration of the contract, and we assess the player's market and performance values during each year of the contract; however, the methods by which we make these predictions and assessments differs between the two outcomes:

- We use the same market model for both prediction and assessment of player value. We predict the average yearly salary for a given player using his performance up to the season preceding the new contract. We assess his market value during each year of the contract using his performance up to and including a given year (e.g. the first year of the contract). For this latter evaluation, the player's market value is equivalent to predicting his average yearly salary if he were to sign a new contract for the subsequent season(s).

- We use the performance model to predict the average yearly value for a new contract, based on translating a prediction of next year's WAR—treated as the average WAR for each year of the new contract for simplicity (future refinements will relax this simplification)—into dollars using the position-specific conversion factors summarized in Appendix A. As an example, one WAR produced by a utility player is worth more than any other position player, whereas one WAR for a closer has the highest value for pitchers. These conversion factors were calculated directly from our dataset using the same subset of players used to fit the models. This approach differs from other WAR-based evaluations that use a fixed conversion factor (typically on the order of $4–7M). Swartz has performed multiple studies on the cost of a win [10,11], but we opt to use our data to determine these values because our study covers a different time period and contains different players and position breakdowns than previous analysis. Similar to Swartz's analysis, our conversions show significant variability in the cost of a win for players who play different positions.

We employ several supervised machine-learning algorithms to train the models. These algorithms include linear regression, linear regression with feature selection—including forward selection, backward elimination, correlation feature selection, feature selection based on random forest variable importance, and regularized regression via elastic nets—regression trees, and gradient-boosted trees. We select the best performing algorithm for each outcome based on the cross-validated root mean squared error between the predicted and observed outcomes.

We define predicted surplus value (pSV) and surplus value (SV) to facilitate comparison between the model predictions and player salaries. The pSV is the difference between the predicted value and the observed new salary, whereas SV is the difference between the assessed value and the observed new salary. A positive pSV suggests that players were undervalued (or their agent did not negotiate well), whereas a negative pSV indicates that players were overvalued. A positive SV suggests that players outperformed their salary, whereas a negative SV suggests that a player underperformed. Values close to zero for pSV suggest a close match between a player's negotiated salary and the predicted value of his performance, whereas values close to zero for SV suggest the same for a player's salary and the assessed value of his actual performance.

We compute two additional measures based on pSV and SV. The first measure is the correlation coefficient r between pSV and SV, where each pSV and SV is aggregated over the duration of each free agent contract. The second measure is consistency (ξ), which is the percentage of player contracts with pSV and SV having the same sign. These two measures attempt to capture the degree to which each model consistently evaluates players, that is, players who are predicted to perform below the expectations set by their free agent contracts actually underperform, and players who are predicted to perform above the expectations actually perform well.

4. RESULTS

We present the results of our analysis in three subsections. In the first subsection, we summarize the results of the supervised machine-learning algorithms for the market and performance models for batters and pitchers. In the second subsection, we present our analysis of free agent performance and the ability of each model to predict and assess player value. In the final subsection, we present the results of a retrospective study that uses both models to identify high-value free agents.

4.1. Model Quality and Interpretations

The results of each supervised machine-learning algorithm are summarized in Appendix B. The models' performance on batters and pitchers models was quite similar, with the market models explaining a higher proportion of the dependent variable variation than the performance models. It is worth noting that the dependent variables for these two approaches are quite different; therefore, we cannot directly compare the measures of fit. In general, the linear regression models with feature selection—most notably the stepwise selection algorithms—performed significantly better than the full linear regression models or tree-based models (i.e., regression tree, gradient-boosted trees), which tended to overfit the data. Models fit to the log-transformed dependent variable (i.e., average yearly salary) performed relatively poorly for the market model (after translating the dependent variable back to the original units). Based on these results (and for simplicity), we select the backward elimination regression model for each scenario for the remainder of our analyses as the results are indistinguishable from the results for the forward selection models.

Despite our exclusion of the other algorithms for our remaining analysis, we can still glean many useful insights about how player performance affects the estimates for yearly average salary and future WAR. In particular, the statistical significance of features in linear regression models, the features that survive feature selection approaches, and feature importance derived from tree-based methods all provide insight about the most meaningful player attributes and statistics. We describe some of the most insightful interpretations about the models in the paragraphs that follow.

The most important feature (by far) for the market model was the player's current salary. Current salary achieved statistical significance for all models, survived every feature selection algorithm, and ranked first with respect to feature importance for all tree-based methods. All else being held constant, players who were well paid in the past were likely to continue to be well paid in the future. In a sense, the current salary of a given player is a proxy for all of the information used to determine the duration and value of that player's previous contract. We describe additional statistically significant features and insights (in no particular order) below:

- Age: All else held constant, younger players signing free agent contracts tended to be paid more than older players.

- Year prior to contract signing: Yearly free agent markets differ substantially, and similar performances may be rewarded differently based on the supply and demand of players during a particular offseason.

- Players re-signed by current teams: All else being held constant, players who re-signed with their current teams were paid more on average. In general, better players—who warrant higher salaries—were more likely to stay with their current teams. This trend may reflect that—from the team's perspective—these players are more difficult to replace (either via free agency or with in-house talent) than lesser players, and therefore, there is an incentive to retain these players even if they are more expensive.

- Team market size: We represent this factor most directly by team payroll but also by team attendance. Payroll was a much more important factor than attendance, which suggests that large-market teams pay higher salaries. However, in addition, teams with high attendance may feel more pressure to sign better players in order to satisfy a dedicated fan base.

- Advanced performance metrics: These metrics tended to be significantly more important than traditional

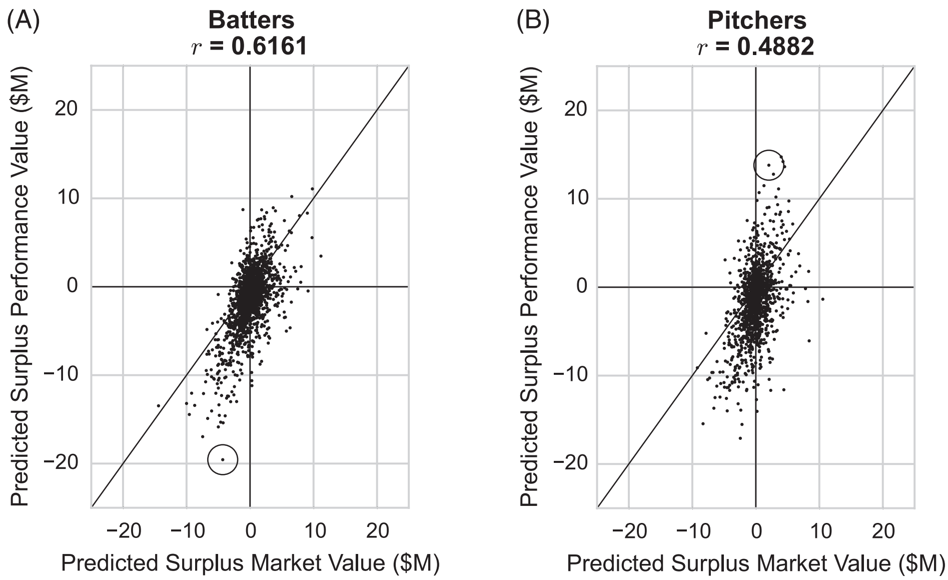

Fig. 2 (a) Scatter plots of predicted surplus values for the market versus the performance models for batters (left) and pitchers (right), including highlighted predictions for Ryan Howard (circle, left) and Chris Sale (circle, right). (b) Scatter plots of predicted surplus values for the market versus the performance models for batters (left) and pitchers (right), including highlighted predictions for Ryan Howard (circle, left) and Chris Sale (circle, right).

performance metrics. Batters with high WAR (particularly offensive WAR), runs created, and win probability added as well as pitchers with high game scores, WAR, win probability added, and rates of quality starts and strikeouts were paid substantially higher. More traditional stats such as batting average, slugging percentage, hits, home runs, and stolen bases rated much lower for batters as did wins, hits allowed, and earned run average for pitchers. These results are a testament to the recent development of more meaningful performance measures by Sabermetricians.

- Career statistics: Career statistics (cumulative and average) play an important role, but to a lesser degree than the aforementioned features. Players who performed poorly during a particular season could be compensated well based on accomplishments in previous seasons.

- Playoff statistics: Playoff statistics played a fairly small role overall but were more important for pitchers than batters. Many players do not have significant playoff experience, which may explain the limited role of playoff statistics in the models.

The performance models tend to be much simpler than the market models, that is, they include far fewer meaningful features. As one would expect, the most important feature for all players for the performance model is WAR (i.e., in the year preceding the new contract). All else held constant, players who generated high WAR—including high offensive WAR for batters—were expected to continue to perform well in the near future. For batters, power–speed (i.e., the ability to hit home runs *and* steal bases) and the ability to draw walks (either intentionally or at a high rate) were also important predictors. For pitchers, the ability to strike out batters, limit base runners, consistently pitch deep into games, and contribute to team wins were all important. Career stats again played a small role, and playoff achievements had no impact.

4.2. Free Agent Performance and Model Evaluation

We utilize the best market and performance models for batters and pitchers (i.e., the backwards elimination regression) to predict average yearly salary for a new free agent contract and assess player value during each contract. In Fig. 2, we plot pSV for the market model against pSV for the performance model for batters and pitchers, respectively. These figures clearly show that the two models agree on pSV for many players (across a broad range of pSV values), as indicated by the moderately strong positive correlations. However, there are a significant number of players for whom the two models disagree quite strongly. For example, the market model predicted an average yearly surplus value of -$4.3M for the Philadelphia Phillies' first baseman Ryan Howard (relative to his 5-year/$125M contract) after the 2011 season, whereas the

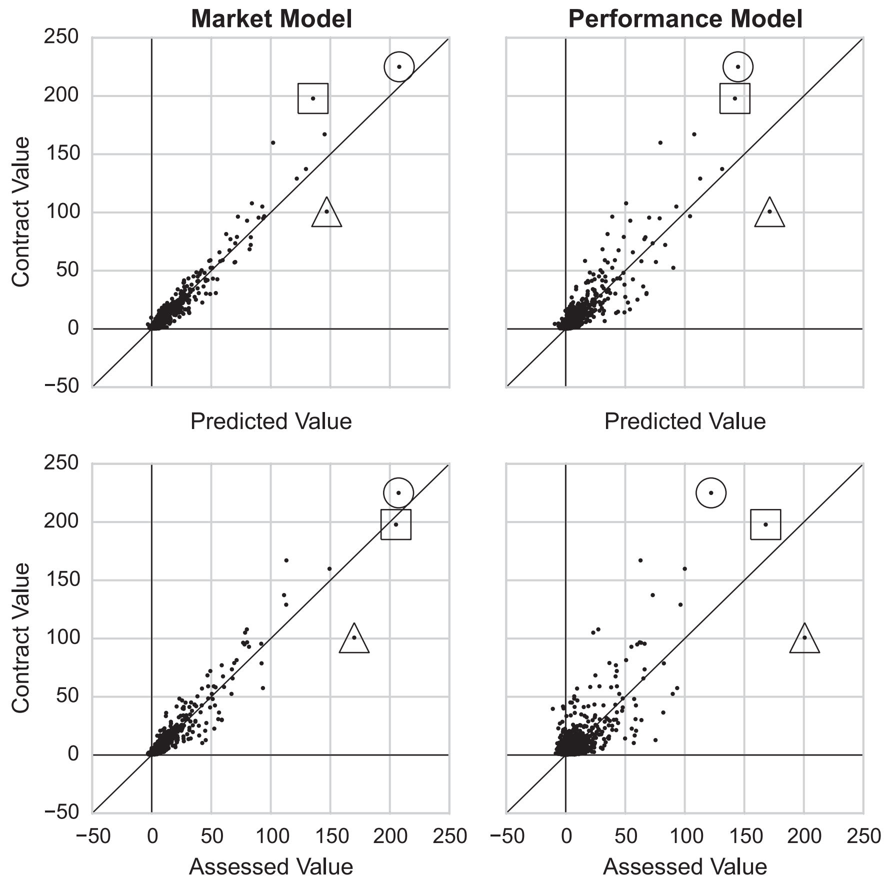

Fig. 3 Comparison of predicted and assessed values for the batter market and performance models, including highlighted contracts for Derek Jeter (2001, circle), Alex Rodriguez (2001, square), and Albert Pujols (2004, triangle).

performance model predicted an average yearly surplus value of −19.6M (as highlighted in Fig. 2). As many would agree (subjectively or objectively), Howard has performed quite poorly since signing this particular contract. As another example, the market model predicted an average yearly surplus value of $2.1M for the White Sox starting pitcher Chris Sale (relative to his 5-year/$32M contract) after the 2012 season, whereas the performance model predicted an average yearly surplus value of $13.8M (as highlighted in Fig. 2). Sale has become one of the best starting pitchers in the MLB.

Next, we compare the cumulative predicted and assessed values for the two models for batter and pitcher contracts, respectively, in Figs. 3 and 4. The first apparent trend from these figures is that—relative to the contracts—there is less variation of the predicted and assessed values for the market model, which reflects that the market model, as expected, more closely predicts contract values. By contrast, the performance model demonstrates more variation and thus offers a different perspective on whether players have under- or over-performed the expectations set by their new contracts.

We highlight several examples in these figures to demonstrate the differences between the two models. For the batters, in Fig. 3, we highlight Derek Jeter's 10-year/$189M contract signed in 2001 (circle), Alex Rodriguez's 10-year/$252M contract signed in 2001 (shortened to approximately 7-years/$158M after opting out after the 2007 season, square), and Albert Pujols's 7-year/$100M contract signed in 2004 (triangle). These players received three of the largest contracts in MLB history and produced three of the highest evaluated performances for any contract in our dataset.

The contracts for Rodriguez and Jeter have been two of the most scrutinized free agent contracts in MLB history, and both models predicted that these two players were paid more than their value. The models generally agree on Rodriguez, predicting that his contract was approximately $50M too high. For Jeter, however, there is a more than $50M disagreement, with the market model evaluating Jeter at an amount much closer to his actual contract value. The models disagree on the assessed value of the two players. The market model assessed that the two players produced values very close to their contracts, whereas the

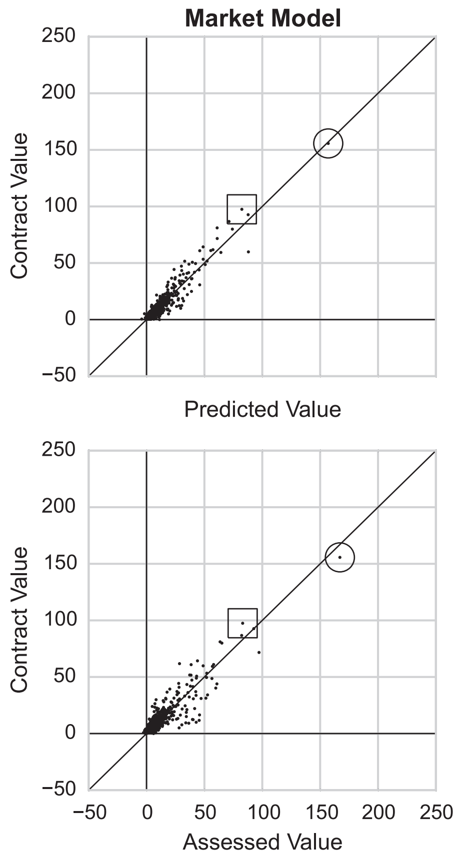

Fig. 4 Comparison of predicted and assessed values for the pitcher market and performance models, including highlighted contracts for Randy Johnson (1999, circle), and Carlos Zambrano (2008, square).

performance model assessed that they fell significantly short, with Jeter being assessed a SV of approximately −$100M. Pujols's contract represents an example for which both models agree that his predicted and assessed values far exceeded that of his contract. The predicted and assessed values for both models placed Pujols's value between $150M and $200M over the course of his 7-year contract, which far exceeds his contract value. In the case of the performance model, Pujols produced more value over 7 years than Rodriguez over the same period of time and Jeter over 3 additional years.

For the pitchers (see Fig. 4), we highlight Randy Johnson, who's contract we inferred to be 9 years and $122M in the absence of complete contract information. In actuality, this inferred contract consists of a 4-year/$52M contract signed with the Arizona Diamondbacks in 1999, a player option for the 5th year, a 2-year contract extension signed with the New York Yankees after he was traded from Arizona in 2005, and an unknown 2-year contract. Despite the inaccuracy of this inferred contract, Johnson proves to be an interesting example. In addition, we highlight Carlos Zambrano (square), who signed a 5-year/$91M contract with the Chicago Cubs that went into effect for the 2008 season. The market model for both players predicted and assessed values approximately equal to their contract values. However, the performance model valuated Johnson at over $50M more than his expensive inferred contract, whereas Zambrano fell more than $50M short.

We have provided some examples of how the market and performance models predict and assess player value. As stated previously, the market model is much more accurate at predicting player salaries for new contracts. By contrast, the performance model highlights players whose assessed values stray significantly from their contract values. Both models show another interesting trend, which is that it is unusual for players to produce more than $50M of assessed value over the course of a given contract and extremely rare for players to produce more than $100M of the same. As a result, there is a clear trend that players signed to multiyear contracts are more likely to be overpaid, which has been explored by other researchers [12,13]. This trend is supported by the observation that Pujols and Johnson—who had extraordinary runs of production exceeding the values of their contracts—are the only players beyond $100M in assessed value that fall below the 45° line in Figs. 3 and 4. This is a strong indication that

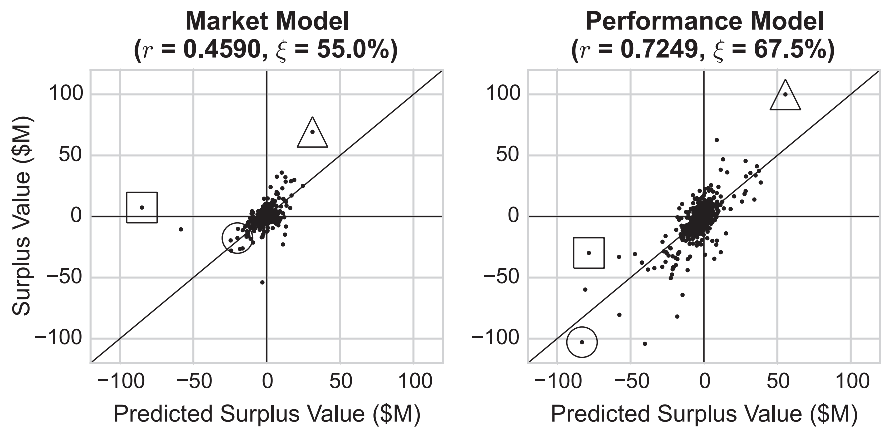

Fig. 5 Comparison of predicted and assessed surplus values for the market and performance models for batters, including highlighted contracts for Derek Jeter (2001, circle), Alex Rodriguez (2001, square), and Albert Pujols (2004, triangle).

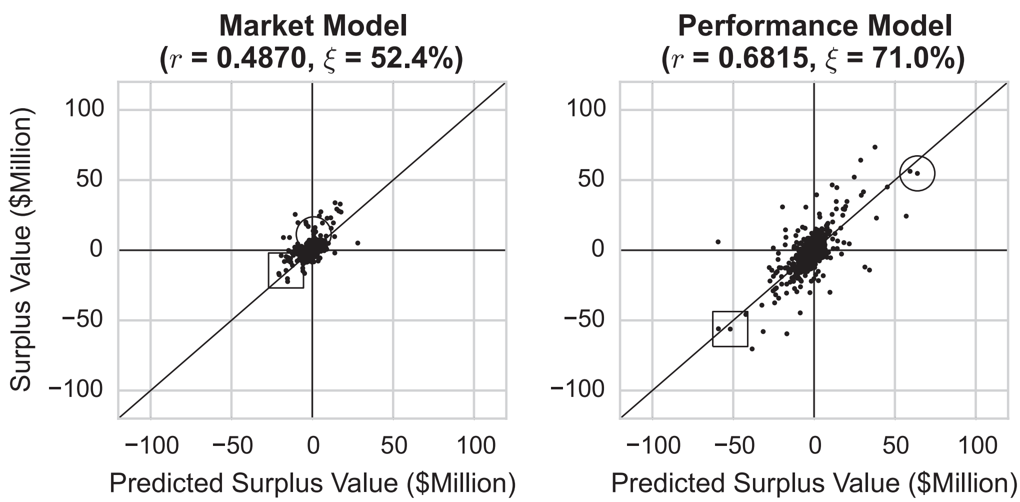

Fig. 6 Comparison of predicted and assessed surplus values for the market and performance models for batters, including highlighted contracts for Randy Johnson (1999, circle), and Carlos Zambrano (2008, square).

so-called ‘mega contracts,’ such as the recent 13-year/$325 contract given to outfielder Giancarlo Stanton by the Miami Marlins [14], often result in disappointment for the teams that award such contracts as well as their fans.

An alternate viewpoint is to consider player performance on a yearly basis, which we summarize for batters and pitchers in Appendix C. With these figures, we are able to compare player performance between contracts of different lengths rather than the cumulative viewpoint taken in Figs. 3 and 4. We observe very similar trends with respect to the market and performance models for the aforementioned examples of Jeter, Rodriguez, Pujols, Johnson, and Zambrano, who are still outliers in terms of their salaries. However, there are other players who produced exceptional average assessed values (e.g., greater than $15M/year), albeit over the course of shorter contracts.

Lastly, we compare pSV to SV for batter and pitcher contracts for each model in Figs. 5 and 6. We observe that the predictions for Jeter, Johnson, and Zambrano are aligned well with their assessments. However, Rodriguez and Pujols exceeded their predictions significantly, with the former being overly pessimistic and the latter being insufficiently optimistic. For all players, we observe that the performance model has a higher correlation between pSV and SV and a higher rate of consistency. In other words, a player’s predicted value correlates more closely with his assessed value for the performance model than for the market model, which provides support for the performance model as being more reliable for identifying players to pursue and players to avoid. In addition, the performance model appears to be able to identify players who fall towards both ends of the performance spectrum, that is, players who perform really well and really poorly compared to their contracts. For additional examples, we list the best and worst contracts for batters and pitchers—as assessed by SV—in Appendix D.

Table 1. Summary of batters with the highest excess predicted surplus value with respect to the performance model. All salaries and values are reported in units of one million 2014 U.S. dollars.

Player	Pos	Year	Age	Team	WAR	Predicted WAR	Next WAR	New average yearly salary	Performance Value	Market value	Excess value
Ichiro Suzuki	RF	2003	29	SEA	5.6	6.1	9.1	13.9	22.6	12.8	9.8
Carlos Santana	C	2011	25	CLE	3.8	3.7	4	4.3	13	6.2	6.9
Andrew McCutchen	CF	2011	24	PIT	5.7	5.2	7	8.8	15.5	9	6.6
J.D. Drew	RF	2004	28	ATL	8.3	5.1	3.2	11.5	19	12.6	6.4
Elvis Andrus	SS	2011	22	TEX	4.2	4.2	4	4.9	12.6	6.3	6.3
Kurt Suzuki	C	2009	25	OAK	3.4	3	2.2	4.4	10.7	4.5	6.2
David Eckstein	SS	2004	29	ANA	1.9	3.4	2.9	3.5	10.2	4	6.1
Andrelton Simmons	SS	2013	23	ATL	6.9	4.6	3.5	8.3	13.7	7.7	6
Carl Crawford	LF	2004	22	TBD	4.9	3.8	4.4	4.6	12.3	6.3	6
Jose Reyes	SS	2006	23	NYM	5.8	5.1	5.1	6.7	15.1	9.3	5.8

Table 2. Summary of pitchers with the highest excess predicted surplus value with respect to the performance model. All salaries and values are reported in units of one million 2014 U.S. dollars.

Player	Pos	Year	Age	Team	WAR	Predicted WAR	Next WAR	New average yearly salary	Performance value	Market value	Excess value
Chris Sale	SP	2012	23	CHW	5.9	5.5	6.9	6.6	20.4	8.7	11.7
Barry Zito	SP	2001	23	OAK	4.5	4.8	7.2	3.1	17.8	7.1	10.7
David Wells	SP	2003	40	NYY	4.3	3.5	2.6	1.6	13.1	2.9	10.2
Tim Lincecum	SP	2009	25	SFG	7.5	6.3	3.7	12.5	23.2	13	10.2
Joakim Soria	CL	2008	24	KCR	3.7	3.4	2.7	3.2	16	6	10
James Shields	SP	2007	25	TBD	5.5	4.7	3.8	3.1	17.3	7.4	9.9
Johan Santana	SP	2004	25	MIN	8.6	5.9	7.2	12.1	21.9	12.3	9.6
Clayton Kershaw	SP	2011	23	LAD	6.5	6.4	6.2	9.8	23.4	14.3	9.1
Aaron Harang	SP	2006	28	CIN	5.2	4.9	6	10.5	18	9.1	8.9
Josh Johnson	SP	2009	25	FLA	6.6	4.3	7.2	10.6	15.8	7.9	8

4.3. Retrospective Study of High-Value Free Agents

To further evaluate the potential of the performance model as an aid for decision making, we perform a retrospective study of impending MLB free agents, based on players with the largest differences between pSV for the two models (i.e., the largest vertical distances from the 45° line in Fig. 2). In Tables 1 and 2, we identify batters and pitchers with the highest excess predicted value, that is, the largest positive difference between pSV for the performance model and pSV for the market model. These are players who were predicted to generate large assessed values for the performance model relative to their actual contracts (i.e., players predicted to fall toward the upper right of the performance model plots in Figs. 5 and 6). In other words, they represented opportunities for teams to sign players to less expensive contracts relative to their

Table 3. Summary of batters with the lowest excess predicted surplus value with respect to the performance model. All salaries and values are reported in units of one million 2014 U.S. dollars.

Player	Pos	Year	Age	Team	*WAR*	Predicted WAR	Next WAR	New average yearly salary	Performance value	Market value	Excess value
Ryan Howard	1B	2011	31	PHI	1.2	1.9	−1.2	25.8	6.2	21.5	−15.3
Derek Jeter	SS	2010	36	NYY	1.7	2.3	1.1	17.9	6.7	18.2	−11.5
Manny Ramirez	LF	2007	35	BOS	1.1	1.7	6	20.9	5.5	16.7	−11.2
Miguel Tejada	SS	2009	35	HOU	1.9	0.6	0.6	6.5	1.9	12.8	−11
Barry Bonds	LF	2006	41	SFG	4	2.2	3.4	17.8	7.2	18.1	−10.9
Pat Burrell	LF	2007	30	PHI	1.5	0.8	2.3	15.7	2.7	13.6	−10.9
Magglio Ordonez	RF	2009	35	DET	0.8	0.9	1.3	19.4	3.5	14.4	−10.9
Preston Wilson	CF	2005	30	COL, WSN	−0.4	−0.1	−0.9	4.7	−0.3	10.4	−10.7
Richie Sexson	1B	2007	32	SEA	−1.1	0.5	−0.1	17.1	1.8	12.5	−10.7
Ryan Howard	1B	2008	28	PHI	1.8	1.6	3.8	20	5.4	16.1	−10.7

Table 4. Summary of pitchers with the lowest excess predicted surplus value with respect to the performance model. All salaries and values are reported in units of one million 2014 U.S. dollars.

Player	Pos	Year	Age	Team	WAR	Predicted WAR	Next WAR	New average yearly salary	Performance value	Market value	Excess value
A.J. Burnett	SP	2013	36	PIT	1.7	−0.1	0.1	11.2	−0.4	15.7	−16
Andy Ashby	SP	2002	34	LAD	1.4	−1.7	−0.4	11	−6.1	8.5	−14.6
Mike Hampton	SP	2008	35	ATL	0.1	−1	−0.6	2.2	−3.8	10.6	−14.5
Bronson Arroyo	SP	2013	36	CIN	2.5	−0.6	0.7	11.8	−2.3	11.9	−14.2
Tim Lincecum	SP	2013	29	SFG	−0.6	1.6	−0.7	17.5	5.8	19.8	−14
Kevin Brown	SP	2001	36	LAD	3.1	1.4	−0.4	20.7	5.2	18.5	−13.3
Hiroki Kuroda	SP	2013	38	NYY	4.1	0.9	2.4	16	3.3	16.1	−12.8
Al Leiter	SP	2004	38	NYM	4.8	−0.5	−0.9	8.5	−1.9	10.6	−12.5
Chan Ho Park	SP	2006	33	SDP	−0.4	−0.2	−0.3	0.7	−0.7	11.2	−11.9
Greg Maddux	SP	2002	36	ATL	4.4	1.4	1.3	19	5.1	16.7	−11.5

future performance levels. In Tables 3 and 4, we list players who were highly overvalued by the market relative to their predicted future performance. These are players who could have been identified as high-risk acquisitions. In these tables, we also provide information on the WAR of the player for the season preceding the new contract, the average future WAR predicted by the performance model, and the player's actual WAR for the subsequent season. This information provides some indication about how the player performed relative to the predictions.

There are some interesting results from this approach of identifying players with excess predicted surplus value. First, most of the players listed in Tables 1 and 2 appear to be younger players who were on the verge of breakout seasons and/or careers. These players would be worth taking some additional risks to acquire in the free agent (or trade) market as they have the potential to be productive for several years in the future. In the case of Tables 3 and 4, almost all the players were older and past their peak performance years and were paid quite well based on previous levels of success. However, they were unable, at older ages, to produce at levels of performance matched by their contracts. Therefore, this approach appears to be quite effective at identifying players to avoid in the free agent market.

5. CONCLUSIONS

We have presented results from models trained to predict and assess the value of MLB free agents. The market models demonstrated a reliable ability to evaluate players according to observed practices. The performance models, although less reliable than the market models in terms of their predictive accuracy, demonstrated the ability to reliably identify players who were both highly overvalued (e.g., Ryan Howard in 2011) and highly undervalued (e.g., Chris Sale in 2012). In addition, the performance model could be used to identify future stars as well as established stars who are not likely to be as productive in the future.

APPENDIX A

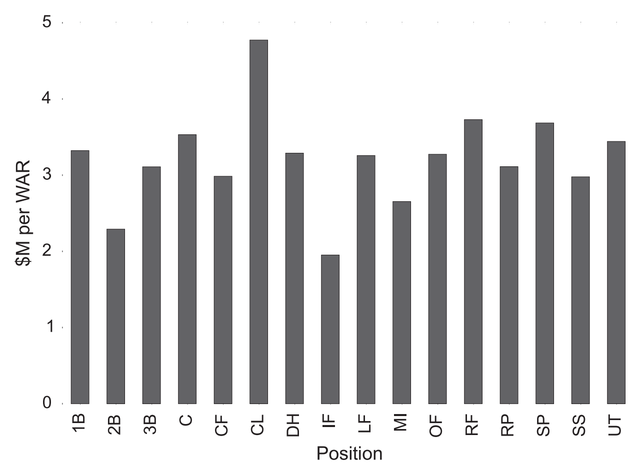

Fig. A1 2014 US dollars per wins above replacement (WAR) by position, aggregated across the free agent players in the data set used for the analysis. The figure shows that a closer (CL) earns the most on average per WAR for pitchers, whereas relief pitchers (RP) earn the least. For batters, utility players (UT) earn the most on average per WAR, whereas infielders (IF) earn the least. Other positions in the figure include: first base (1B), second base (2B), third base (3B), catcher (C), center field (CF), designated hitter (DH), left field (LF), middle infielder (MI), outfield (OF), right field (RF), starting pitcher (SP), and shortstop (SS).

APPENDIX B. MODEL PERFORMANCE TABLES

The following tables summarize the performance measures of the algorithms used. The predictions of the log models are translated back to actual dollars before performance measure calculations.

Table B1. Market model training and cross-validation results for batters.

Model	Training R^2	Cross-validated R^2
Backward elimination (Linear)	0.7963	0.7721
Forward selection (Linear)	0.7894	0.7709
Elastic net (Linear)	0.7710	0.7573
Random forest feature selection (Linear)	0.7649	0.7503
Linear regression (Linear)	0.8026	0.7455
Gradient-boosted trees (Linear)	0.8322	0.7344
Gradient-boosted trees (Log)	0.8283	0.7087
Correlated feature selection (Linear)	0.6848	0.6808
Regression tree (Linear)	0.8136	0.6308
Forward selection (Log)	0.6726	0.6150
Regression tree (Log)	0.7822	0.6146
Elastic net (Log)	0.5981	0.6072
Random forest feature selection (Log)	0.6525	0.6005
Backward elimination (Log)	0.6641	0.5638
Correlated feature selection (Log)	0.5839	0.5566
Linear regression (Log)	0.6856	0.4640

Table B2. Performance model training and cross-validation results for batters.

Model	Training R^2	Cross-validated R^2
Backward elimination	0.4091	0.3754
Forward selection	0.3967	0.3754
Random forest feature selection	0.3654	0.3547
Elastic net	0.3619	0.3484
Gradient-boosted trees	0.4737	0.3289
Correlated feature selection	0.3284	0.3272
Linear regression	0.4306	0.3093
Regression tree	0.2771	0.2332

Table B3. Market model training and cross-validation results for pitchers.

Model	Training R^2	Cross-validated R^2
Backward elimination (Linear)	0.8029	0.7522
Forward selection (Linear)	0.7921	0.7511
Gradient-boosted trees (Linear)	0.8459	0.7427
Random forest feature selection (Linear)	0.7540	0.7296
Elastic net (Linear)	0.7658	0.7292
Gradient-boosted trees (Log)	0.7895	0.7231
Elastic net (Log)	0.7577	0.7127
Correlated feature selection (Linear)	0.7142	0.7001
Forward selection (Log)	0.7363	0.6973
Random forest feature selection (Log)	0.7170	0.6858
Linear regression (Linear)	0.8128	0.6817
Backward elimination (Log)	0.7378	0.6670
Regression tree (Linear)	0.8354	0.6330
Regression tree (Log)	0.8022	0.5956
Correlated feature selection (Log)	0.5967	0.5615
Linear regression (Log)	0.7521	0.5141

Table B4. Performance model training and cross-validation results for pitchers.

Model	Training R^2	Cross-validated R^2
Backward elimination	0.4428	0.3593
Forward selection	0.3977	0.3593
Random forest feature selection	0.3307	0.2896
Elastic net	0.2119	0.2873
Gradient-boosted trees	0.4341	0.2824
Correlated feature selection	0.2772	0.2584
Linear regression	0.4790	0.2194
Regression tree	0.2836	0.2173

APPENDIX C

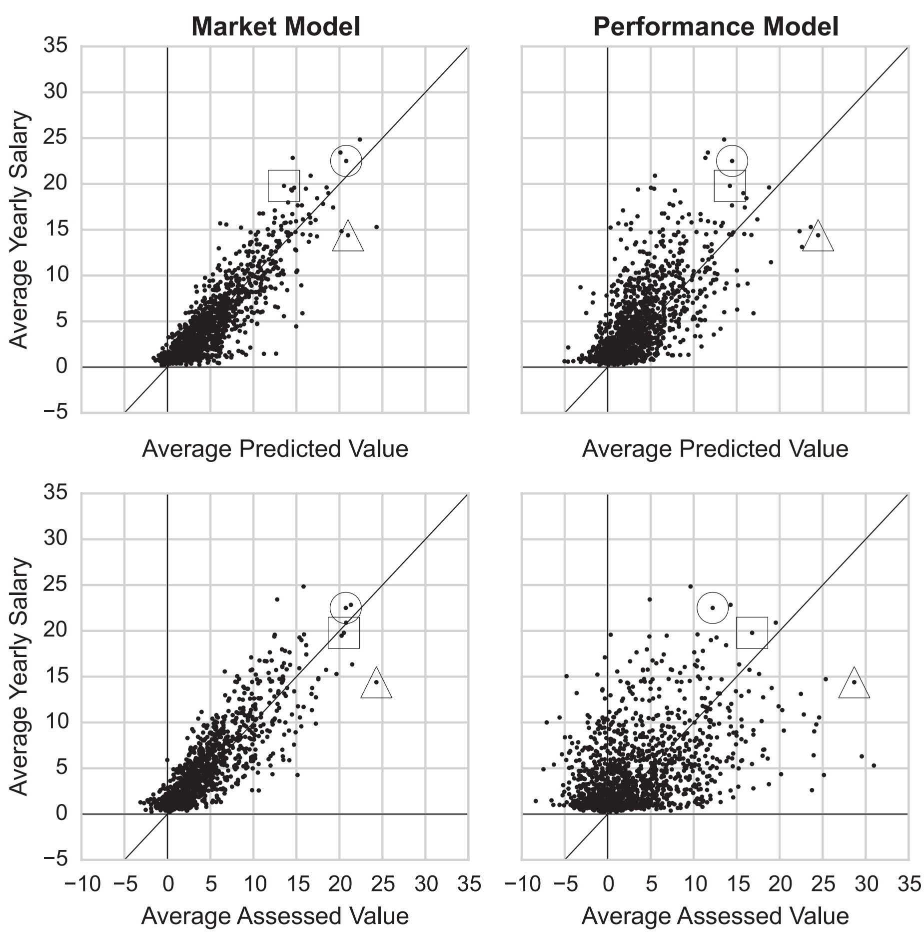

Fig. C1 Comparison of average predicted and assessed values for the batter market and performance models. The predicted value is the prediction of a player's value based on his past performance (i.e., in advance of a new contract) whereas the assessed value is the value of a player's performance during each year of a given contract, according to each respective model. The market model (left column) predicts new average salary for a given player-season, whereas the performance model (right column) predicts WAR for a given player-season and translates this prediction into value using the average cost of a WAR per position. We highlight contracts for Derek Jeter (2001, circle), Alex Rodriguez (2001, square), and Albert Pujols (2004, triangle). For example, the market model valued Albert Pujols more highly than his newly signed contract in 2004 (upper left), and his market value throughout his contract also exceeded what he was paid (lower left). The performance model tells the same story (i.e., that he was undervalued), as his predicted and assessed performance value was significantly higher than his actual contract (upper and lower right, respectively).

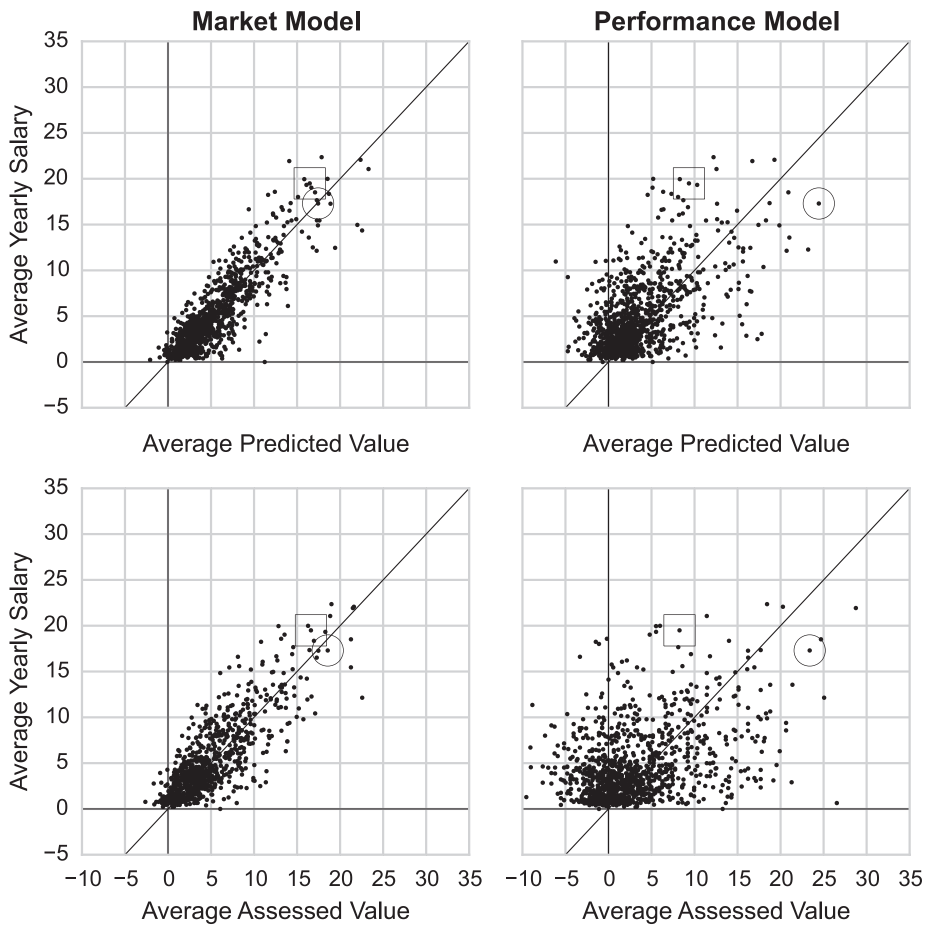

Fig. C2 Comparison of predicted and assessed values for the pitcher market and performance models. The predicted value is the prediction of a player's value based on his past performance (i.e., in advance of a new contract) whereas the assessed value is the value of a player's performance during each year of a given contract, according to each respective model. The market model (left column) predicts new average salary for a given player-season, whereas the performance model (right column) predicts WAR for a given playerseason and translates this prediction into value using the average cost of a WAR per position. We highlight contracts for Randy Johnson (1999, circle) and Carlos Zambrano (2008, square). According to the market model, the predicted (upper left) and assessed values (lower left) for both Randy Johnson and Carlos Zambrano were very close to the values of their new contracts, as well as each other's predicted and assessed values. However, the performance model predicted that Johnson would perform quite well relative to his contract, whereas Zambrano's performance would fall short of his contract (upper right). It turns out that the performance model predictions were consistent with Johnson's and Zambrano's actual performance (lower right), which match well with more subjective evaluations of their performance over these time periods.

APPENDIX D. BEST AND WORST CONTRACTS FOR BATTERS AND PITCHERS

Table D1. Best contracts for batters according to the market and performance models. The best contracts are identified as those with the largest positive difference (surplus value, SV) between their actual salary and the assessed value of their performance. All salaries and values are reported in units of one million 2014 U.S. dollars.

Market model						Performance model					
Player	Year signed	Age	Duration	Salary	SV	Player	Year signed	Age	Duration	Salary	SV
Albert Pujols	2004	24	7	100.8	69.3	Albert Pujols	2004	24	7	100.8	100
David Wright	2007	24	6	57.5	35.9	Jason Giambi	1999	28	3	12.8	62.6
Jason Giambi	1999	28	3	12.8	32.6	Adrian Gonzalez	2007	25	4	10.3	46.8
Adrian Gonzalez	2007	25	4	10.3	32.2	Joe Mauer	2007	24	4	36.5	45.5
Ian Kinsler	2008	26	5	22.6	29.9	Victor Martinez	2005	26	5	16.8	41.2
Curtis Granderson	2008	27	5	30	28.9	Vernon Wells	2003	24	5	20.8	38.3
Robinson Cano	2008	25	4	30.3	28.6	Ichiro Suzuki	2004	30	4	52.5	37.4
Brian McCann	2007	23	6	30.8	25.1	David Wright	2007	24	6	57.5	36.1
Vernon Wells	2003	24	5	20.8	24.1	Carl Crawford	2005	23	4	14.2	35.6
Victor Martinez	2005	26	5	16.8	23.1	Curtis Granderson	2008	27	5	30	33.6

Table D2. Worst contracts for batters according to the market model and performance models. The worst contracts are identified as those with the largest negative difference (surplus value, SV) between their actual salary and the assessed value of their performance. All salaries and values are reported in units of one million 2014 U.S. dollars.

Market model						Performance model					
Player	Year signed	Age	Duration	Salary	SV	Player	Year signed	Age	Duration	Salary	SV
Alex Rodriguez	2008	32	10	167.1	−54	Alex Rodriguez	2008	32	10	167.1	−104.2
Carlos Lee	2007	31	6	107.9	−27.9	Derek Jeter	2001	27	10	225	−102.9
Alfonso Soriano	2007	31	8	105.1	−26.6	Alfonso Soriano	2007	31	8	105.1	−81.9
Jason Giambi	2002	31	7	137.3	−26.1	Carlos Lee	2007	31	6	107.9	−80.5
Gary Sheffield	2004	35	5	72.1	−22.9	Jason Giambi	2002	31	7	137.3	−64.2
Manny Ramirez	2009	37	2	46.9	−21.3	Manny Ramirez	2001	29	7	159.9	−59.8
Torii Hunter	2008	32	5	96.4	−19.6	Jose Guillen	2008	32	3	39.6	−50.5
J.D. Drew	2007	31	5	77.1	−18.2	Jorge Posada	2008	36	4	57.1	−47.5
Derek Jeter	2001	27	10	225	−17.7	Gary Matthews	2007	32	4	41.7	−44
Ichiro Suzuki	2008	34	5	95	−17	Carlos Delgado	2005	33	5	68.4	−43.5

Table D3. Best contracts for pitchers according to the market and performance models. The best contracts are identified as those with the largest positive difference (surplus value, SV) between their actual salary and the assessed value of their performance. All salaries and values are reported in units of one million 2014 U.S. dollars.

Market model						Performance model					
Player	Year signed	Age	Duration	Salary	SV	Player	Year signed	Age	Duration	Salary	SV
Tim Hudson	2001	25	4	11.6	33.8	Tim Hudson	2001	25	4	11.6	73.5
James Shields	2008	26	4	10	32.9	Bartolo Colon	1999	26	4	12.5	64.2
Jake Peavy	2005	24	4	16.5	29.3	Barry Zito	2002	24	4	12.3	56.3
Barry Zito	2002	24	4	12.3	28	Randy Johnson	1999	35	9	155.7	54.8
Adam Wainwright	2008	26	4	9.1	27.4	Matt Cain	2007	22	4	5	52.1
Bartolo Colon	1999	26	4	12.5	27.2	Adam Wainwright	2008	26	4	9.1	46.5
John Smoltz	2002	35	6	71.7	25.5	Jake Peavy	2005	24	4	16.5	45.1
Ubaldo Jimenez	2009	25	4	9.5	24.5	Dan Haren	2006	25	4	7.7	44.6
Matt Cain	2007	22	4	5	23.1	Johan Santana	2005	26	4	31	41.6
Dan Haren	2006	25	4	7.7	22.8	Doug Davis	2003	27	6	21.7	39.5

Table D4. Worst contracts for pitchers according to the market and performance models. The worst contracts are identified as those with the largest negative difference (surplus value, SV) between their actual salary and the assessed value of their performance. All salaries and values are reported in units of one million 2014 U.S. dollars.

Market model						Performance model					
Player	Year signed	Age	Duration	Salary	SV	Player	Year signed	Age	Duration	Salary	SV
Bartolo Colon	2004	31	4	60.9	−22.3	Mike Hampton	2001	28	5	81	−70.4
Derek Lowe	2009	36	4	64.3	−20.2	Chan Ho Park	2002	29	4	61.8	−59.6
Mike Hampton	2001	28	5	81	−17.5	Derek Lowe	2009	36	4	64.3	−58
Jeff Suppan	2007	32	4	44.3	−16.5	Carlos Zambrano	2008	27	5	97.5	−56.2
Chan Ho Park	2002	29	4	61.8	−16.5	Kevin Brown	2002	37	4	79.9	−56
Kevin Brown	2002	37	4	79.9	−14.8	Jeff Suppan	2007	32	4	44.3	−45.8
Carlos Zambrano	2008	27	5	97.5	−14.4	Bartolo Colon	2004	31	4	60.9	−44.7
Tom Glavine	2003	37	3	40.9	−12.8	Carlos Silva	2008	29	3	36.6	−39.2
John Lackey	2010	31	2	37.1	−10.9	John Lackey	2010	31	2	37.1	−37.5
Eric Milton	2005	29	3	29.9	−9.8	Oliver Perez	2009	27	3	26.4	−34.2

REFERENCES

[1] J. K. Hakes and R. D. Sauer, An economic evaluation of the Moneyball hypothesis, J Econ Perspect 20 (3) (2006), 173–185.

[2] A. Farrar and T. H. Bruggink, A new test of the Moneyball hypothesis, Sport J 14 (1) (2011), 1–9.

[3] M. Lewis, Moneyball: The Art of Winning an Unfair Game, New York, NY, W.W. Norton & Company, 2003.

[4] D. W. Rockerbie, Strategic free agency in baseball, J Sports Econ 10 (3) (2009), 278–291.

[5] J. Turvey, The future of baseball contracts: a look at the growing trend in long-term contracts, Baseball Res J 42 (2) (2013).

[6] Baseball-Reference.com., 2015. URL http://www.baseball-reference.com.

[7] ESPN MLB Free-Agent Tracker, 2015. URL http://espn.go.com/mlb/freeagents.

[8] Cot's Baseball Contracts, 2015. URL http://www.baseball prospectus.com/compensation/cots/.

[9] RETROSHEET Transactions, 2015. URL http://www.retrosheet.org/transactions.

[10] M. Swartz, A retrospective look at the price of a win, part two, 2012. URL http://www.fangraphs.com/blogs/a-retrospective-look-at-the-price-of-a-win-part-two/.

[11] M. Swartz, Positional differences in the price of WAR, 2012. URL http://www.fangraphs.com/blogs/positional-differences-in-the-price-of-war-2/.

[12] A. C. Krautmann, and M. Oppenheimer, Contract length and the return to performance in Major League baseball, J Sports Econ 3 (6) (2002), 6–17.

[13] K. Stankiewicz, Length of contracts and the effect on performance of MLB players, Park Place Econ 17 (2009), 76–83.

[14] B. Nightingale, Marlins, Giancarlo Stanton finalize 13-year, $325 million deal, 2014, URL http://www.usatoday.com/story/sports/mlb/2014/11/14/giancarlo-stanton-marlins-contract-extension/19040675/.