7TH ANNUAL
MIT SLOAN
SPORTS ANALYTICS CONFERENCE
MARCH 1-2, 2013 BOSTON CONVENTION AND EXHIBITION CENTER
presented by ESPN

Does a 'coattail effect' influence the valuation of players in the Major League Baseball draft?

Kevin Whitaker
Princeton University
Princeton, NJ 08544
Email: kswhitak@princeton.edu

Abstract

This paper introduces the "coattail effect" in the Major League Baseball amateur draft, in which top college baseball prospects draw substantial attention from professional scouts, who become more likely to see the star player's teammates and more likely to recommend selecting them with later draft picks. Using a sample of 11,540 college players drafted in 1984-2003, strong evidence is found that more players are drafted from a given college team when there is a star teammate present. Estimated effects of the coattail effect on the value players provide to major-league teams are mixed and not very conclusive.

1 Introduction

The spring of 2011 signified a landmark season for the University of Connecticut baseball team. Connecticut won its first Big East Conference regular-season championship in team history and reached the Super Regional round of the NCAA Tournament, and another accomplishment took place in June, when 10 members of the team were selected in the Major League Baseball amateur draft. Before 2011, an average of 1.7 Connecticut players were drafted per year, and never had more than six been selected in the same season.

What made 2011 such an exception? Naturally, the main factor was that the team had many good players, which is why it performed so well in the college season. But Connecticut's talent was not unprecedented – for example, its 1979 team reached the College World Series, yet only two players were drafted that summer. What set Connecticut apart in 2011 may have been the presence of George Springer and Matt Barnes, who were great enough to be selected with two of the top 20 overall picks of the draft. In the MLB draft, the value each team can expect to receive from a pick declines significantly in the later rounds [1]. Some overlooked players become important big leaguers, but such success stories are the exception and not the rule: from 1984-2003, 66.4 percent of first-round picks reached the major leagues, but only 8 percent of players drafted later than the tenth round did. Therefore, teams spend most of their time and resources scouting and evaluating the very top amateurs – players such as Springer and Barnes – to optimize their earliest draft picks.

Because most late-round picks will not reach the major leagues and because there is relatively little variation between the talent levels of amateurs who would be reasonably considered for a late-round pick, teams and scouts spend comparatively few resources optimizing their later selections. (In one example, the Texas Rangers used their 33rd-round pick of the 2011 draft on a player who had been permanently paralyzed during a college game that spring, purely as a goodwill gesture.) Teams instead are said to rely on certain heuristics to draft players in the later rounds, including familiarity. According to this theory, a scout is more likely to recommend a player he has seen more frequently – perhaps because he is a teammate of a first-round pick, who the scout has made a point to see several times – than a player of comparable talent with whom he is less familiar.

Anecdotal evidence confirmed that the Connecticut baseball team was seen by more scouts than usual in 2011 [2], and a scouting director for the Boston Red Sox believed that the extra exposure from playing with Barnes and Springer helped more teammates be drafted by pro teams [3]. The history of draft picks from the University of Connecticut lends some supporting evidence to this theory: An average of 2.75 teammates were drafted when at least one member of the team was drafted in the first five rounds, compared to 1.38 total players in other years, though much of the effect comes from the last three drafts.

But this effect is certainly not limited to UConn – it should exist wherever a top prospect is bringing scouts' eyeballs to his less-talented teammates, causing noticeable effects in the valuation of late-round MLB draft picks. Even when good decisions can result in a multi-million-dollar return from a very small investment, are highly seen players systematically chosen more frequently? And, if so, are there more "hidden gems" among less familiar players?

2013 Research Paper Competition
Presented by:
42 ANALYTICS
SAP

This paper describes that theory as a "coattail effect." The term is adapted from political science, in which a coattail effect occurs when a strong presidential candidate for one party improves the chances that candidates from the same party will be elected for lower offices. This is broadly similar to the baseball phenomenon studied in this paper: In political elections, a strong presidential candidate attracts more voters who support his party, giving them the opportunity to also vote for his partisan "teammates"; in baseball, a star prospect attracts more scouts to see his team, giving them the opportunity to evaluate his teammates and consider them with draft picks. The "coattail effect" has been used in political literature for several decades [4], but as of now, neither it nor any other term has consistently been used to describe baseball scouting.

This analysis focuses only on players who were drafted from four-year colleges. The coattail theory is invoked more frequently about college players than high school players, and the information regarding college attendance is more reliable than high school data. Additionally, because there are so many more high school baseball players who play for so many more high schools, the base rate of players drafted from each high school is very small, making it more difficult to find evidence of players riding coattails of top prospects. Finally, much scouting at the high-school level is done at showcase events, in which the best amateurs form select youth teams to compete against each other in tournaments specifically designed to draw the attention of scouts; this process would not result in the same coattail effect that is expected to manifest itself in college players.

2 Methodology and Results

Part 1: Measuring the coattail effect

The first model looks for evidence of a coattail effect – that teammates of a top prospect are more likely to be drafted in later rounds than players who do not have an attention-attracting teammate. But there are clearly biases present when trying to answer this question – top prospects are more likely to play for good baseball programs, and other good players are also more likely to play for good programs. A simple model would almost certainly show that having a top teammate is associated with being drafted, but this result would be biased by the omitted variable of program strength.

To isolate the influence of a "top prospect," which is defined here as a player drafted in the top five rounds, a panel data regression model is used, including a fixed-effects parameter for the college a player attended.

Within the fixed-effects regression, the number of non-top prospects (players selected after the fifth round) chosen from that school in the years when a star teammate was drafted can be compared to the number of non-top prospects drafted in other years. If there is a "coattail effect" in the draft, there should be a correlation between the presence of a star prospect and the number of non-stars drafted, when accounting for the school the players were drafted from.

That is, the regression takes the form:

$$D_{it} = \beta_0 + \beta_1 * Coattail_{it} + a_i + u_{it} \qquad (1)$$

where D_{it} is the number of players drafted from a school i in year t in rounds 6+, $Coattail_{it}$ is a dummy variable equal to 1 if a player from that school was drafted in the first five rounds in that year and 0 if not, a_i is a fixed effect for each school on the number of players that will be drafted in an average year, β_1 is the effect of a top prospect on the number of players drafted, and u_{it} is the residual.

It is also possible that the coattail effect extends into other years. For example, if a player is one of the top prospects in year Y, scouts might also see players who will not be draft-eligible until year Y+1 and remember them when the time comes, so the coattail effect might continue into the following year. It is also plausible that scouts will want to see the top player in year Y-1 (to get an earlier evaluation of his talent), and in doing so, they will see players who are draft-eligible that year, meaning the coattail effect would extend to the previous year (though, intuitively, this seems less likely). To measure these, the model is expanded to include indicator variables for the year before and after a top prospect was drafted:

$$D_{it} = \beta_0 + \beta_1 * Coattail_{it} + \beta_2 * Prev_{it} + \beta_3 * Next_{it} + a_i + u_{it} \qquad (2)$$

This analysis included all 11,540 college players, representing 890 different schools, who were drafted from 1984-2003. The number of players drafted after the fifth round were tabulated and grouped by college and year, leaving a tally of the number of players who were drafted from each of 17,800 school-years (890 colleges x 20 years), as well as a dummy variable indicating the presence of a star teammate who was drafted that year.

On average, 2.66 players were drafted from a given school in seasons with a possible coattail effect – seasons in which a top prospect was drafted from the same school – while only .43 players per school were drafted in other seasons. This simple difference is biased, however, because the list of top prospects is dominated by players who were drafted from baseball powerhouses, which naturally produce more players who are good enough to be drafted. When tested with a regression that includes fixed effects for each school, the coattail effect remains present and positive, but it is substantially smaller in magnitude. When controlling for these effects, a top prospect's coattails are associated with an increase of .76 players drafted in a given season (t=23, p<<.001).

Table 1. Panel regression of all players drafted after fifth round, 1984-2003

	1	1a	2
	# Drafted	# Drafted	# Drafted
coattail	0.7631**	-61.858**	0.7484**
	(.0329)	(9.6422)	(.0326)
year		-0.0004	
		(.0012)	
coattail*year		0.0314**	
		(.0048)	
prev			0.3803**
			(.0332)
next			0.5190**
			(.0331)
school	*fixed effects*	*fixed effects*	*fixed effects*
constant	0.5216**	1.3088	0.4707**
	(.0069)	(2.361)	(.0073)

***, *, † represent significance at 1, 5 and 10% levels, respectively.*

The coattail effect has appeared to grow stronger over time – the estimate of an interaction term between coattail and year was positive and highly significant, as seen in regression 1a (t=6.49, p<.001). The coattail effect also seems to carry strongly into years before and after a top player was drafted. Indicator variables for the year before and the year after a player was drafted in the first five rounds were both highly significant, as seen in regression 2. Each category was associated with an increase of about .4-.5 players drafted per year (p<<.001 in all cases).

There remain other possible explanations for these effects that go beyond scouting attention and a coattail effect. As above, it is reasonable to think that the existence of a top prospect in one year might convince more good (draft-caliber) players to join the program the following years, which could explain the magnitude of *next*; similarly, it is reasonable to think that the existence of more good (draft-caliber) players in a program might convince a top prospect to join the program in a following year, which could explain the magnitude of *previous*. Additionally, though this method controls for school-specific effects, it holds those effects fixed over time. If a college team improves significantly in quality over time, it might have top prospects, as well as draft-caliber players, in later years than in earlier years; such a bias would not be accounted for by this model.

Part 2: Value of draft picks and the coattail effect

From the statistical evidence examined in previous MLB drafts, which supports the theories and anecdotal evidence, it appears that the coattail effect is a significant factor in explaining the decisions that teams make when selecting players in later rounds of the draft. In theory, the only attribute that clubs should try to optimize in these situations is the expected value of drafting each player, which is a direct function of each player's true talent level and should not be related to the coattail effect.

This raises the possibility that teams' reliance on the coattail effect causes them to make less optimal draft selections. Imagine if there was a public list of all draft-eligible players, ranking each by their true talent level. In this case, if teams were swayed by a coattail effect and were more likely to draft teammates of top prospects, they would be making suboptimal decisions – they would not always be choosing the most talented player available, and would therefore not be maximizing their eventual major-league value. (One note: in this context, 'true talent' refers to not just physical abilities but to everything that might make a player become a valuable major-league contributor, such as work ethic, knowledge of the game, etc.)

Of course, such a list is not possible in reality; scouts have to rely on imperfect signals of true talent, and the difference between the talent levels of low draft picks is often undiscernibly small. Thus, it is also plausible that the coattail heuristic does not affect the returns teams receive on draft picks.

The value of a drafted player is measured in two separate ways. In the first method, the dependent variable is a binary outcome of whether or not the player reaches the major leagues; in the second, the regressed variable is his career Wins Above Replacement, calculated by Baseball-Reference.com. The two models have relative advantages and disadvantages – by treating all major-league careers as if they were identical, the former does not use some available information regarding how valuable a player was to his team; the latter, on the other hand, is susceptible to large outliers, as one star player could have the career value of 100 lesser major-league players, as well as small sample sizes, because these players make the major leagues infrequently. This is consistent with the existing literature, which is largely split between models that use the probability of reaching the majors and those that use measures of major-league performance.

To judge how scouts value players, one additional piece of information is used – the round in which a player was drafted. This acts as a proxy for how scouts judge a given player's talent – teams should always select the player they believe is best among the remaining pool in each round, so a player drafted in a later round is expected to be less valuable than any player drafted before him.

A player's expected chance of reaching the major leagues is estimated with the following model, using all players drafted in rounds 6+ from 1984-2003:

$$\Pr(MLB_i) = a_0 + a_1{*}Round_i + a_2{*}(Round_i)^2 + a_3{*}(Round_i)^3 + a_4{*}Coattail_i + u_i \quad (3)$$

where $\Pr(MLB_i)$ is the probability that a player i reaches the major leagues, a_0 is a constant term, $Round_i$ is the round in which a player was drafted, $Coattail_i$ is a dummy variable indicating whether or not a player had a teammate who was a top prospect, a_4 is the effect of playing with a top prospect on the chance that a player reaches the major leagues, given the round he was drafted in, and u_i is the residual. The model is expected to be cubic in *Round* based on the work of Spurr [5], who found statistically significant coefficients for the second and third degrees of a player's draft position on the probability that he reached the major leagues.

A player's expected performance if he reaches the major leagues, also expected to be cubic with respect to *Round*:

$$\ln(WAR_i) = \beta_0 + \beta_1{*}Round_i + \beta_2{*}(Round_i)^2 + \beta_3{*}(Round_i)^3 + \beta_4{*}Coattail_i + \varepsilon_i \quad (4)$$

where WAR_i is a player's expected Wins Above Replacement, β_0 is a constant term, β_1 is the percentage effect of a one-round increase on a player's expected WAR, β_4 is the effect of playing with a top prospect on a player's expected WAR, given the round he was drafted in, and ε_i is the residual. A log-linear model is used to reduce the effect of large outliers in WAR, as the distribution of Wins Above Replacement is skewed heavily to the right by a handful of major-league stars.

If the coattail effect does not irrationally change how scouts evaluate amateur players, the "coattail" variable should have no effect on the probability that a player reaches the major leagues or on his major-league production, so the coefficients a_4 and β_4 would not be significantly different from zero. But if there is a systematic bias towards selecting familiar players, and if this leads to an irrational over-valuation of teammates of top prospects, those coefficients should be negative – players that scouts are familiar with, who are valued as equal to overlooked players (without a top prospect), as evidenced by the round in which they were selected, should be less likely to offer value in the major leagues.

This model is based on the assumption that teams always select the player they feel to be the best available prospect at that point in the draft. The main reason this may not necessarily be true is that some players have large bonus demands – in other words, the expected cost of signing them is higher. In that case, a player who is expected to provide somewhat less value to a major-league team but will cost less to sign may give the team more expected surplus value, so he will be drafted more highly. This happens most frequently with high school or junior college players, who have more leverage in negotiations, because they are younger and have the alternative of going to college and returning to the draft later. By studying only players from four-year colleges, the majority of these players are left out of the sample. In an ideal world, the models would also control for the signing bonus a player received (or demanded, in the case that he did not sign); unfortunately, teams generally prefer not to disclose the bonuses they pay – and in cases in which the player does not sign, the counterfactual value of how much money it would have taken to convince the player to sign is impossible to determine – so this data is not very accessible, particularly for later rounds and for older drafts. This will only bias my models if there are a significant number of college players who fall into this category and if the propensity to do so is correlated with the variable of interest, whether or not a player benefited from coattails; I consider the second assumption, in particular, unlikely.

Regressing the return on draft picks by round and coattails can show whether or not any variation in value is explained by the coattail effect. Only players who were drafted from rounds 6+ were included, as the performance of other players is not relevant to the study of the coattail effect. The first model uses whether or not a player reached the major leagues as the dependent variable.

Contrary to the theoretical prediction, players who benefit from the coattail effect do not perform worse than their draft position-based expectation – they have actually been more likely to reach the major leagues, as seen in regression 5 (z=2.12, p=.034). A probit model provides a somewhat better fit for the data than a linear model, but the basic conclusions drawn from each model are the same.

Table 2. Professional success of college players drafted after fifth round, 1984-2003

	3a	3b	4
	(Linear) Pr(MLB)	(Linear, demeaned) Pr(MLB)	ln(WAR)
coattail	0.0142*	-0.0176*	-0.135
	(.0068)	(.0083)	(.1537)
round	-0.0218**	-.0209**	0.0455
	(.0018)	(.0019)	(.0627)
$round^2$	.00046**	.00043**	-0.0034
	(.00005)	(.00005)	(.0025)
$round^3$	$-2.9*10^{-6}$**	$-2.7*10^{-6}$**	0.00005
	$(4.7*10^{-7})$	$(5.0*10^{-7})$	(.00003)
school		*absorbed*	
constant	0.3734**	0.3706**	0.5506
	(.0168)	(.0176)	(.4273)
(Pseudo) R^2	0.0451	0.0321	0.0122

***, *, † represent significance at 1, 5 and 10% levels, respectively.*
Pr(MLB) is the estimated probability a player will play a game in the major leagues;
WAR is Wins Above Replacement, only included for players who played at least one game in the majors.

The positive effect of *coattail* on the chance that a player will reach the major leagues, given his draft position, disappears when school-fixed effects are accounted for in regression 3b; in fact, the estimated coefficient becomes significantly negative, instead of significantly positive, at the 5% level. It is imprudent to conclude simply that scouts overvalue players based on the coattail effect, because of the results in regression 3a – if coattails influence the value of draft picks, that effect should exist between schools, not just within them. However, regression 3b gives some supporting evidence that the presence of a top prospect may cause scouts to overvalue his teammates, compared to how they would normally value other players from that school. Because this result is not extremely significant in a large sample, and because it occurs under a very precise definition, it should be treated with some skepticism.

In regression 4, a coattail effect was not a significant predictor of value produced among players that did reach the major leagues, as measured by a log transformation of Wins Above Replacement. Notably, none of the *round* variables were significant predictors of ln(*WAR*) either, which is surprising, given that better players should consistently be selected earlier in the draft. (This remains true for other functional forms of this model, including regressions for an untransformed value of *WAR* and regressions with only a first-order round variable.) For college players drafted in the sixth round or later who make the major leagues, it appears that their MLB value is essentially unpredictable from any of these variables.

3 Conclusions

In the Major League Baseball draft, teams make six- or seven-figure investments on young prospects that may or may not become players of major-league caliber, buying proverbial lottery tickets that can yield extreme benefits to baseball teams. With an abundance of data that is meticulously recorded and a major economic incentive to maximize investments, there should not be systematic biases that interfere with this value-maximization. But previous research has already shown that some of these biases in player valuation do exist.

This paper gives evidence for the existence of another heuristic that teams use in evaluating amateurs, and specifically players worthy of late-round draft picks: The coattail effect. When a college baseball team has a star prospect – a player good enough to draw the attention of large numbers of scouts – his teammates are more likely to be drafted in later rounds. Anecdotal evidence suggests that, in later rounds, scouts are more likely to recommend players with whom they are familiar, which is consistent with the findings in this study; the relation holds for not only the year in which a top prospect is drafted, but also for the previous year and the two following years. On average, when controlling for the average differences between college teams, an additional .76 players were drafted from a school in a season in which that team had a top prospect. Given that an average of only .43 players were drafted per year in seasons without a top prospect, this represents a very substantial increase in the number of teammates expected to receive a chance to play professionally. This effect also seems to have become stronger over time in the sample investigated by this study, spanning 1984-2003. (Further studies could investigate if this trend in the coattail effect has continued since 2003, although it is too early at this point to make any conclusions about how much value teams received from more recent draft selections.)

If scouts do rely on this coattail effect, it is theoretically possible that they are making suboptimal decisions, as their goal is to always select the most talented player available with each draft pick, regardless of the quality of their teammates. However, there is currently little evidence in support of that theory. Players who were teammates of a top prospect actually over-performed their expected performance, given the round in which they were selected, when measured by how likely they were to eventually reach the major leagues. A model using major-league performance, conditional on reaching the major leagues, showed no significant predictive value from a variable indicating the presence of a coattail effect.

The "coattail effect" in baseball scouting is very difficult to detect in available data, just as it is in politics [6]. The attention given to various players by scouts is not directly measurable, so there are alternative explanations that are consistent with the patterns observed in this study. Most notably, the talent of other players on a team may not be independent of whether or not there is a star player on the roster. Most college players that will be draft-eligible in a given year are recruited over a long, concurrent process, so there is not likely to be a strong relationship between the talent level of players within a certain recruiting class (holding constant the program for which they will play); still, there may be cases when a player commits to a program at least partially on the basis of another player's concurrent commitment. In this way, it is plausible that a star player's presumed commitment will cause more draft-quality players to commit as well, which might cause more players than usual to be drafted along with the top prospect.

An extension of this research, not possible with the dataset used in this study, which might account for this bias would be to look at only the number of seniors who were drafted in the same year as a top prospect who is a junior; the seniors' college choice could not be influenced by the junior's, isolating the coattail effect. This would, however, leave open the possibility of causality in the opposite direction – a star prospect is more likely to join a team with more talented players. In an ideal study, players worthy of becoming top prospects would select college teams based on exogenous factors; but this clearly does not happen in practice, as college team selection for all players is a dynamic process that depends partly on the caliber of their potential teammates.

4 Acknowledgements

I would like to thank my adviser, Prof. Andriy Norets, and Rohit Lamba for their advice and guidance in conducing this research.

5 References

[1] Stephen J. Spurr, "The Baseball Draft: A Study of the Ability to Find Talent," Journal of Sports Economics, vol. 1, no. 1, pp. 66-85, Feb. 2000.
[2] Greg Auman, "UConn Deals with Scouting Scrutiny," Baseball America – College Baseball, Feb. 20, 2011.
[3] Viv Bernstein, "With a Harvest of Prospects, UConn Baseball Makes Its Mark," The New York Times, June 13, 2011.
[4] Malcolm C. Moos, Politics, Presidents and Coattails, Greenwood Press, London, 1952.
[5] Stephen J. Spurr, "The Baseball Draft: A Study of the Ability to Find Talent," Journal of Sports Economics, vol. 1, no. 1, pp. 66-85, Feb. 2000.
[6] Warren E. Miller, "Presidential Coattails: A Study in Political Myth and Methodology," Public Opinion Quarterly, vol. 19, no. 4, pp. 353-68, Winter 1955.