Omega 90 (2020) 102001

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Data analytics effects in major league baseball☆

Ramy Elitzur

Rotman School of Management, University of Toronto, Toronto, Ontario, M5S-3E6, Canada

ARTICLE INFO

Article history:
Received 5 January 2018
Accepted 9 November 2018
Available online 14 November 2018

Keywords:
Data analytics
Moneyball
OR in sports
Empirical analysis
Information
Major league baseball

ABSTRACT

The use of data analytics has enjoyed resurgence over the last two decades in professional sports, businesses, and the government. This resurgence is attributable to Moneyball, which exposed readers to the use of advanced baseball analytics by the Oakland Athletics, and how it has resulted in improved player selection and game management. Moreover, it changed managerial vocabulary, as the term "Moneyballing" now commonly describes organizations that use data analytics. The first research question that this study examines is whether the organizational knowledge related to baseball data analytics has provided any advantage in the competitive Major League Baseball (MLB) marketplace. The second research question is whether this strategic advantage can be sustained once this proprietary organizational knowledge becomes public. First, I identify "Moneyball" teams and executives, i.e., those who rely on baseball data analytics, and track their pay/performance over time. Next, using econometric models, I analyze whether these "Moneyball" teams and GMs, have enjoyed a pay-performance advantage over the rest of MLB, and whether this advantage persists after the information becomes public.

1. Introduction

Data analytics is defined as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact based management to drive decisions and actions" [13]. One of the early (and most famous) applications of data analytics is the use of advanced baseball statistics, or Sabermetrics, by Billy Beane and the Oakland Athletics to excel in the competitive major league baseball (MLB) marketplace, as documented in the best-selling book Moneyball [35]. The book resulted in the resurgence of the use of data analytics tools. The importance of the book is further demonstrated by the fact that the term "Moneyball" has become part of our daily vocabulary. Organizations that use data analytics are often described as playing "Moneyball", or "Moneyballing" [42].

The impact of Moneyball [35] and its popularity go well beyond sports culture. While the book sold over a million copies, it was subsequently made into a movie in 2011. The movie starred Brad Pitt as Billy Beane, and was nominated for six Academy Awards, including Best Picture and Best Actor. Ironically, before the release of the movie, the book entered into popular culture, shown by the fact that the Simpsons had an episode about it in 2010 titled MoneyBart [1].

The book's importance for operations research (OR) is evidenced by the fact that data analytics and related OR tools, as a result of Moneyball [35], have been enthusiastically embraced by both businesses and governmental departments and agencies [12,27,42,43]. Moreover, the importance of data analytics for management science as a discipline is demonstrated by the fact that *Management Science* [6] and *Omega* [16], have dedicated special issues to the use of data analytics in business. Furthermore, *Interfaces*, has dedicated an issue to the use of data analytics specifically in sports [25,26].

This study contributes to the existing literature by providing insights on data analytics as proprietary organizational knowledge in the competitive MLB marketplace, and on whether the resulting strategic advantage is absolute or comparative and diminishes once it becomes public. Baseball is uniquely suitable for the study of the effects of data analytics as one can identify the precise point of time in which advanced analytics were adopted by teams. Therefore, one can examine the overall effects on the MLB afterwards. Data analytics is especially suitable for decision-making in baseball organizations because the game is less team-oriented than other major league sports (e.g., the NFL, NBA, or the NHL). While baseball has a two-player interaction (a pitcher confronting a batter) other major league sports have greater intra-team interaction making statistical analysis much more complex. Analyzing the informational advantage provided by data analytics in a competitive marketplace, and whether it is absolute or comparative, is operationalized using the concepts of the Adaptive Market Hypothesis (AMH) in the context of the MLB.

The remainder of the paper is organized as follows. First, the literature review is provided. Next, the background information and

☆ This manuscript was processed by Associate Editor M. Fry.

E-mail address: elitzu@rotman.utoronto.ca

https://doi.org/10.1016/j.omega.2018.11.010
0305-0483/

the research hypotheses are discussed. This is followed by a discussion of the sample and data used in the study. The methodology section follows the data section and covers the models and empirical tests employed. The results, including robustness tests, are the penultimate portion of the paper. Lastly, concluding remarks are offered, including the limitations of the study, and avenues for future research.

2. Literature Review

Lo [38–40] proposes the theory of adaptive market efficiency in capital markets, which maintains that market dynamics are driven by the interaction among selfish individuals, competition, adaptation, natural selection, and environmental conditions. Lo [39] argues that such adaptation is driven by competition, as the interactions among various market participants are governed by natural selection. Furthermore, this notion reconciles market efficiency with behavioral economics and documented behavioral biases by investors [51]. The AMH argues that while market participants often make mistakes, they learn from them and adapt their behavior accordingly [38–40]. Consequently, the line of studies on the vanishing over time of mispricing anomalies is essentially an application of the AMH [30,44,47,49].

Liberatore and Luo [36] assert that data analytics goes well beyond mere logical analysis that use analytical methodologies. Instead, they argue that data analytics encompasses analytical processes that make data actionable, leading to insights that can be used for organizational decision-making and problem solving. Libratore and Luo [36] go on and suggest that "OR professionals must acquire or strengthen their technical and managerial skills to succeed in the new analytics environment, moving from being "one-and-done" problem solvers to solution providers". (p.323). Thus, arguing that OR has to adapt and embed data analytics as an integral part of its tools. Consequently, Moneyball [35], in essence, describes how the use of data analytics in baseball, in particular the use of Sabermetrics, has led the Oakland Athletics to superior performance despite the low payroll of the team. As such, the use of advanced baseball analytics by the Oakland Athletics is arguably one of the early application of data analytics [52].

The few academic papers on baseball deal mostly with specific aspects of the game instead of the overall pay-performance concept [7,11,14,15,36]. An early study that examines the pay-performance relation in MLB is Scully [48], but it does so at the individual player level, rather than team level, and, moreover, due to the period that it examines, it does not look at the effects of data analytics. Hakes and Sauer [31] provide an interesting analysis of "Moneyball" effects but the study looks only at offensive measures, as opposed to overall team performance (e.g., WAR) and, moreover, does not track individual baseball executives' effects, or identify analytics oriented teams.

Of particular importance is the study by Troilo et al. [52] who examine reality versus perception in the adoption of business analytics tools to enhance gate receipts revenues by North American professional sports teams. The results demonstrate that the adoption of business analytics by professional sport teams has made managers realize that their revenues are growing. The second finding is that, consistent with the managers' perceptions, revenue growth is driven by business analytics.

Another important study is Chan and Fearing [11], who develop a new and highly-tractable optimization-based approach for evaluating baseball rosters in the presence of uncertainty due to the risk of player injury. Using this approach, the authors calculate the value of flexibility in different contexts and compared these values across teams.

An important line of literature focuses on the role of organizational knowledge for sustained strategic advantage. Barney [2] examines the role of firm resources, including information and knowledge, in generating sustained competitive advantage. A prerequisite for the value provided by these resources is that they are controlled by the firm, so it can formulate and implement strategies to improve its effectiveness and efficiency. As such, these assets have to be rare, valuable and, moreover, inimitable. Barney argues that the three reasons for these resources to be inimitable are: (a) unique historical conditions that enabled the firm to acquire these assets, (b) the link between them and the firm's sustained competitive advantage is not well understood, or (c) they are socially complex.

Building on Barney [2], Liebeskind [37] argues that knowledge must be protected for a firm to create and maintain its competitive advantage. One problem with the protection of knowledge is that, as opposed to tangible assets, its ownership cannot be asserted unambiguously and, thus, cannot be protected at low costs (for example, one cannot build a fence to protect knowledge, as opposed to a building). Another problem is that, in contrast with tangible assets, knowledge is not clearly observable and, hence, its expropriation or illegal imitation cannot be easily detected. Liebeskind [37] asserts that the firm must use a cost-benefit analysis to choose between internalizing this knowledge (i.e., doing its best to keep the 'secret sauce' secret) and legal protection (which, for example, necessitates the disclosure of some crucial detail when applying for a patent).

Gold et al. [29] synthesize the ideas of Barney [2] and Liebeskind [37] and apply them to information systems (IS) as part of knowledge process capabilities. The analysis of the responses to the survey that the study employs confirm the importance of knowledge protection in the context of IS–related sustained competitive advantage.

Consequently, I argue that MLB provides a unique opportunity to investigate organizational knowledge and its protection for sustained competitive advantage in the context of major league baseball. Thus, this study contributes to the extant literature by examining "Moneyball" in the context of the following research questions:

1. Does organizational knowledge in the form of data analytics provide a strategic advantage in the competitive MLB marketplace?
2. Is the strategic advantage that data analytics provides to MLB teams absolute, or comparative, thus, diminishes once this organizational knowledge becomes public?

3. Background and hypothesis development

Major League Baseball (MLB) is the oldest of all North American professional sport leagues. It currently has 30 teams in playing two leagues: 15 teams in the American League (AL) and 15 teams in the National League (NL). Each League has three divisions (East, Central and West), with 5 teams in each one. Each of the MLB teams plays 162 games in the regular season, which takes place from early April to early October that year. The regular season games are normally played in three-game-series (sometimes a two- or four-game series), where each MLB team plays 81 games at home and 81 games on the road (most of these games are played within the league that the team belongs to, but each team also participates in 20 interleague games each season). The win-loss record in the regular season determines which teams from each league play in the post-season, and how they are seeded. The seeding of teams is important because it determines the matchups and the home field advantage, which goes to the higher seeded team. Of all major league sports, baseball is arguably the hardest for teams to make it to the post-season as only a third of MLB teams under the current system make it to the post-season. In contrast, in the NFL twelve out of 32 teams make it to the post-season, and in the NBA and the NHL sixteen teams out of 31 make it to the post-season. As such, the win-loss record is crucial for MLB teams and thus it is essential for them to optimally draft players, trade for better players and, of course, optimally manage each game.

Moneyball [35] tells the story of Billy Beane, the GM of the Oakland Athletics, and his quest to use data analytics for player selection (both drafting them and trading with other teams for them) and game management. The shift of the team from using heuristics to data analytics was necessary as the Athletics' ownership changed in the mid-nineties and the new owners decided to cut the payroll. The successful implementation of data analytics by the Oakland Athletics to find undervalued players is evidenced by the fact that the team made it to the playoffs each year between 2000 and 2003, despite having one of the lowest payrolls in MLB.

The study focuses on WAR, which is widely thought to be the most important Sabermetric statistic in baseball [10] and whose calculation is explained in detail in Appendix 1. WAR was developed, as FanGraphs [20] states, to capture the total contributions of a player to his team above the league's baseline.

Of all major league sports, baseball is uniquely suitable for the study of the pay-performance relation because it is essentially a game of one-on-one matchups between a batter and a pitcher. The one-on-one aspect of the game is important as it allows a better statistical analysis of players themselves vs. other major league sports where team dynamics and interactions among players complicate the matter (e.g., the NFL, NBA, NHL and soccer).

Using major league baseball provides a relatively simple way to measure the intrinsic value of data analytics with respect to players, and to test whether markets quickly adapt to valuation anomalies, thus, addressing the question of whether the advantage that data analytics provides is absolute or diminishes once this information becomes public. One of the differentiating features of this paper, compared with other studies, is that it focuses on how organizations adapt within a single industry. This focus on the organization themselves (rather than market prices overall) makes it possible to examine on a deeper level which organizations adapt more quickly, and what the factors are that determine the adjustment process. Moreover, the focus of this study on key executives in MLB organizations allows me to specifically identify the entrance of the new data analytics baseball executive species who, as Lo [38,40] argues, are better suited to deal with the environment, and thus drive out the existing species (non-data- analytics baseball executives), whose decision making process is maladaptive with respect to the new environment in MLB.

Data analytics in the MLB fits Barney's [2] definition of a firm's resource: it is controlled by MLB teams to formulate and implement strategies to improve their effectiveness and efficiency. Moreover, until the publication of book the link between this knowledge and the teams' sustained competitive advantage was ambiguous. Consistent with Liebeskind [37], the organizational knowledge related to the use of data analytics cannot be patented, or otherwise legally protected, and, thus, "Moneyball" teams had to choose internalization of this knowledge to protect it, as opposed to legal protection. Consequently, my first hypothesis states that having a unique data analytics information set provides a competitive advantage for "Moneyball" teams in the competitive MLB marketplace.

Hypothesis 1. *Proprietary organizational knowledge in the form of data analytics results in a pay-performance competitive advantage for "Moneyball" teams over other MLB teams.*

Once Moneyball [35] was published, the link between data-analytics teams' knowledge and their sustained competitive advantage became unambiguous, making this resource imitable. The manner that "Moneyball" knowledge leaks take is through the movement of front office personnel from one team to another, as forewarned by Liebeskind [37]. As such, the second research hypothesis stems from the idea that the competitive advantage that data analytics provides is comparative and not absolute, and, consistent with AMH, once this information becomes available to other market participants, the ensuing advantage vanishes over time.

Hypothesis 2. *Data analytics related organizational knowledge provides a comparative, and not absolute, strategic advantage to "Moneyball" teams and, hence, the pay-performance advantages to "Moneyball" teams over other teams vanish over time once this knowledge becomes publicly available.*

The advantage derived from baseball data analytics, however, could be related to the executive in charge of the front office (the GM) and his philosophy, rather than the team. Moreover, one can argue that, consistent with Lo [38–40] this migration of decision makers is the main driver of the leakage of organizational knowledge. Consequently, the following hypotheses, similar to Hypotheses 1 and 2 above, are in the context of front office executives, as opposed to teams:

Hypothesis 3. *Proprietary organizational knowledge in the form of data analytics provides a pay-performance advantage for "Moneyball" executives over other executives.*

Hypothesis 4. *Data analytics related organizational knowledge provides a comparative, and not absolute, strategic advantage to MLB executives and, hence,* the *pay-performance advantages to "Moneyball" executives over other executives vanish once this knowledge becomes publicly available.*

4. Methods

Define $Win\%_j$ as the percentage wins and relative payroll for team j and $NormPay_j$, as $\frac{Team\ j\ Payroll}{Average\ Payroll\ in\ MLB\ baseball\ that\ year\ without\ team\ j}$ (all variable definitions are provided in Appendix 2). In essence, this pay measure is the team payroll relative to the league average in that year without that team (normalized payroll). The use of normalized payroll helps avoid any potential MLB salary inflation effects, which have occurred during the period studied while overall average win percentages remains roughly the same.

4.1. Tests for Hypotheses 1 and 2

The three sub-periods that are used in the study to test Hypotheses 1 and 2 above are as follows:

1. 1997–2002: The period preceding the publication of Moneyball [35] starting in 1997, the year that Oakland A's began to use advanced data analytics.
2. 2003–2008: The period immediately following the publication of Moneyball [35].
3. 2009–2013: The most recent period, culminating with the last year in the sample.

Define $SABR_1$, $SABR_2$ and $SABR_3$ as fixed effect dummy variables denoting a "Moneyball" (data analytics) team in period 1, 2, and 3, respectively. To obtain a more institutive interpretation of the *NormPay* effects on *Win%*, define *SNP* as scaled *NormPay* by quintiles (scaled between zero and one). Also, three interaction dummies are created to test the variable effects of "Moneyball" teams in period 1, 2 and 3, respectively, $SNP_j{*}SABR_1$, $SNP_j{*}SABR_2$, and $SNP_j{*}SABR_3$. In addition, to control for a small market team effect a fixed effect dummy variable is created for each period, (SM_1, SM_2, SM_3, respectively) and an interaction variable, $(SNP_j{}^*SM_1, SNP_j{}^*SM_2, SNP_j{}^*SM_3)$.

Next, tests are conducted to analyze whether this effect has changed following the publication of Moneyball [35] by running the following model for all periods. The idea in Model 5 is to test whether the win percentage is affected by the scaled normalized payroll, the team WAR (to account for the use of Sabermetrics), small market effects, and "Moneyball" effects. The anticipated effect on *Win%* is positive for *NormPay, TeamWAR*, and the "Moneyball" teams related variables in period 1 and vanish afterwards, as hypothesized. Furthermore, a small market team effect will be reflected in *SM* and the related interaction variable.

$$\begin{aligned} Win\%_j = {} & \alpha_j + \beta_1 SNP_j + \beta_2 TeamWAR_j + \beta_3 SABR_1 \\ & + \beta_4\left(SNP_j * SABR_1\right) + \beta_5 SM_1 + \beta_6\left(SNP_j * SM_1\right) \\ & + \beta_7 SABR_2 + \beta_8\left(SNP_j * SABR_2\right) + \beta_9 SM_2 + \beta_{10}\left(SNP_j * SM_2\right) \\ & + \beta_{11} SABR_3 + \beta_{12}\left(SNP_j * SABR_3\right) + \beta_{13} SM_3 \\ & + \beta_{14}\left(SNP_j * SM_3\right) + \varepsilon, \ j = 1, \ldots, 30 \end{aligned} \tag{1}$$

Finally, tests are conducted to analyze whether this effect has changed following the publication of Moneyball [35] by running the following models separately for each of the three periods. The anticipated signs for the variables in each model should be the same as for Model 1.

$$\begin{aligned} Win\%_{jt} = {} & \alpha_j + \beta_1 NormPay_{jt} + \beta_2 TeamWAR_{jt} + + \beta_3 SABR_t \\ & + \beta_4\left(NormPay_{jt} * SABR_t\right) + \beta_5 SM_{jt} \\ & + \beta_6\left(NormPay_j * SM_{jt}\right) \\ & + \varepsilon; \ j = 1, \ldots, 30, t = 1, \ldots, 3 \end{aligned} \tag{2}$$

4.2. Tests of Hypotheses 3 and 4

Testing individual GM effects allows us to specifically track the movement of decision makers among teams. As such, we lose in terms of positive identification of "Moneyball" effects but gain other insights about decision makers and their effects. Moreover, one can argue that, consistent with Lo [38–40], this migration of decision makers is the main driver of the leakage of organizational knowledge once the idea of data analytics was "outed". Hence, the movement of GMs should be investigated to corroborate "Moneyball" effects.

First, I examine the effects of individual "Moneyball" GMs, where GM_i is the fixed effect (intercept dummy) for GM_i, and $GM_iD^*NormPay_i$ is the related variable interaction effect:

$$\begin{aligned} Win\%_i = {} & \alpha + \beta_1 NormPay_i + \delta_i GM_i \\ & + \gamma_i (GM_i * NormPay_i) + \varepsilon, i = 1, \ldots, 7 \end{aligned} \tag{3}$$

Next, the following model analyzes the pre-"Moneyball" effects versus post-"Moneyball", where $MBGM_i$ refers to the fixed effect of GM_i after 2002 and $MBGM_i{}^*NormPay_i$ is the related variable interaction effect during this period.

$$\begin{aligned} Win\%_i = {} & \propto + \beta_1 NormPay_i + \beta_2 MBYears + \beta_3 (MBYears * NormPay_i) \\ & + \delta_i GM_i + \gamma_i (GM_i * NormPay_i) + \mu_i MBGM_i \\ & + \rho_i (MBGM_i * NormPay_i) + \varepsilon; i = 1, \ldots, 7 \end{aligned} \tag{4}$$

5. Sample and data

5.1. Data

Payroll information from 1985 to 2013 is extracted from Sean Lahman's Database [34]. The Win% information over the same period is imported from Baseball-Reference.com [3]. FanGraphs [20] is used to obtain the overall WAR for all MLB teams from 1997 to 2013.

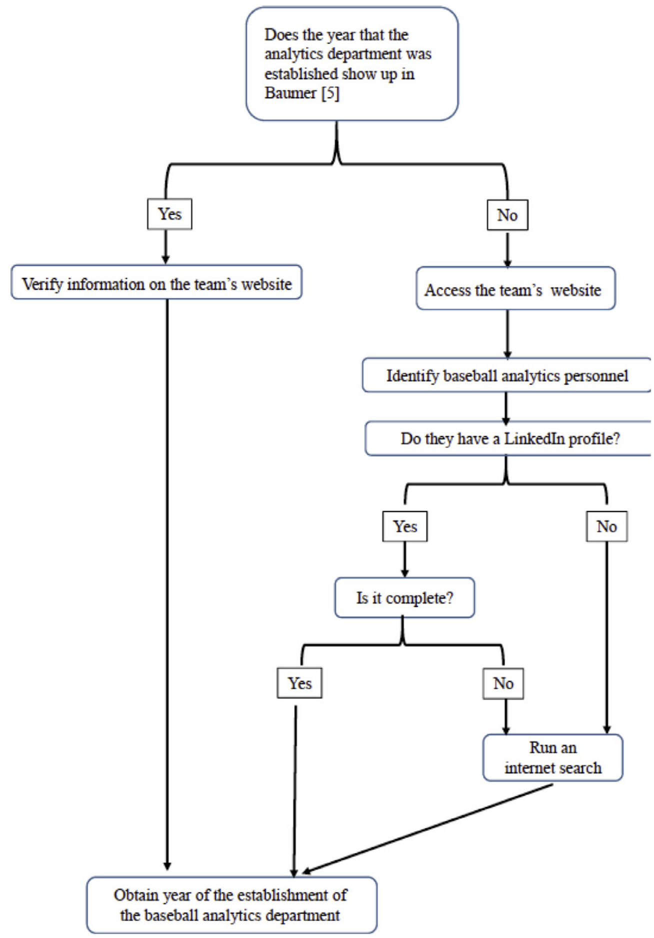

Fig. 1. Structured search diagram.

MLB teams with a baseball operations analytics department are defined to be "Moneyball" teams from the year that this department was established, and are identified in several stages, as depicted in Fig. 1. First, using Baumer [5], information was collected on the year that MLB teams established a data analytics department. Second, to verify the information and complement it, each MLB team's website was accessed to identify data analytics personnel. Third, if the information on Baumer [5], and the team's official website, was not sufficient to determine the year, the LinkedIn profiles of the baseball analytics personnel on the team's official website were analyzed, and the dates were verified through their LinkedIn bios. Fourth, in cases where the LinkedIn profiles were incomplete, missing, or out of date, a rigorous internet search was conducted. Table 1 provides the breakdown of the structured search for the year that the baseball analytics department is established in each MLB team (other than the Oakland Athletics where the information is outlined in Moneyball [35]). Table 2 provides the list of "Moneyball" teams and the years at which the baseball analytics departments were established.

For example, Baumer [5] states in the analysis of the baseball analytics ability of the Baltimore Orioles: "In the office, Sarah Gelles – like Duquette a part of the strong Amherst College pipeline – oversees the O's analytics department, which reaches into both pro scouting and video advance scouting. Gelles built the Orioles' database from scratch, and the team has added Kevin Tenenbaum – a math-econ major who wrote research papers with Dave Allen at Middlebury College – and Pat DiGregory in the past

Table 1
Breakdown of the structured search for the year that the baseball analytics department is established.

Team	Baumer [5]	Team's Website	LinkedIn	Internet Search
Houston Astros	x	x		
St. Louis Cardinals	x	x		
Milwaukee Brewers	x	x		
Seattle Mariners	x	x		
Atlanta Braves	x	x		
Colorado Rockies	x	x		
Detroit Tigers	x	x		
Minnesota Twins	x	x		
Miami Marlins	x	x	x	
Philadelphia Pillies	x	x	x	
Chicago Cubs	x	x	x	
Cleveland Indians	x	x	x	
Pittsburgh Pirates	x	x	x	
Baltimore Orioles	x	x	x	
San Francisco Giants	x	x	x	
Arizona Diamondheads	x	x	x	
Teaxs Rangers	x	x	x	
Boston Red Sox	x	x	x	x
San Diego Padres	x	x	x	x
New York Yankees	x	x		x
Tampa Bay Rays	x	x		x
Kansas City Royals	x	x		x
Los Angeles Dodgers	x	x		x
New York Mets	x	x		x
Toronto Blue Jays	x	x		x
Washington Nationals	x	x		x
Chicago White Sox	x	x		x
Los Angeles Angels	x	x		x
Cincinnati Reds	x	x		x

Table 2
SABR teams over time.

Team	Year analytics are adopted
Oakland Athletics	1997
San Diego Padres*	1997–2001, 2005
Boston Red Sox	2002
San Francisco Giants	2002
Toronto Blue Jays	2002
St. Louis Cardinals	2004
Tampa Bay Rays	2004
Los Angeles Dodgers**	2004–2005, 2015
New York Yankees	2005
Washington Nationals	2006
Kansas City Royals	2006
Cincinnati Reds	2007
Cleveland Indians	2007
Colorado Rockies	2007
Pittsburgh Pirates	2008
Seattle Mariners	2009
Milwaukee Brewers	2010
New York Mets	2011
Texas Rangers	2011
Los Angeles Angels	2012
Baltimore Orioles	2012
Chicago Cubs	2012
Houston Astros	2012
Chicago White Sox	2013
Detroit Tigers	2014
Atlanta Braves	2015
Minnesota Twins	2015
Miami Marlins	2016
Philadelphia Phillies	2016
Arizona Diamondbacks	2017

* The Padres had Theo Epstein between 1997 and 2001 and then started the analytics department in 2005.

** The Dodgers have invested in analytics in 2004-5 but then when Paul DePodesta left have abandoned it until 2014

year." The analysis of Sarah Gelles' LinkedIn profile [28] shows that she was a baseball operations intern in 2009, a labor relations department intern with the MLB in 2010, a baseball operations intern with the Orioles from January to December 2011, a coordinator of baseball analytics for the team from December 2011 to April 2014, a director of analytics from April 2014 to January 2016, and more recently, Director, Analytics & Major League Contracts from April 2014. Based on this information, the Orioles have established a baseball analytics department in the 2012 season, as shown in Table 2.

An example where the fourth step of the identification of Moneyball teams is applied is Theo Epstein, whose LinkedIn profile [18] is out of date (it only shows that he is the GM of the Boston Red Sox while his actual current position is the President of Baseball Operations for the Chicago Cubs). As a result, an internet search was conducted, which shows that he worked for the San Diego Padres as Director of Player Development before joining the Boston Red Sox in 2002 [53]. The internet search [53] also revealed that Theo Epstein, during this time period, studied law at the University of San Diego Law School. Another internet search discovered that he graduated from law school in 2000 [46] and, consequently, working from that date backwards, one can reasonably assert that an analytics department has existed for the San Diego Padres between 1997 and 2001. Baumer [5] states that "2005 and beyond started by Chris Long "With respect to analytics, the Padres have been the class of the NL West for 10 years. Chris Long, a data scientist with an advanced degree in math, started building San Diego's baseball information system in 2004." Analyzing Chris Long's LinkedIn profile [41] reveals that he started in October 2004 as a Senior Quantitative Analyst with the team. Consequently, the Padres re-established their baseball data analytics department in the 2005 season, as shown in Table 2.

Table 2 demonstrates that the number of "Moneyball" teams has grown over time, signaling the proliferation of baseball analytics related organizational knowledge once the book was released.

Table 3
Small market teams by total revenues.

2002		2008		2013	
Team	Revenues ($m)	Team	Revenues ($m)	Team	Revenues ($m)
Montreal Expos	66	Miami Marlins	139	Cleveland Indians	186
Miami Marlins	76	Kansas City Royals	143	Kansas City Royals	169
Kansas City Royals	85	Pittsburgh Pirates	144	Oakland Athletics	173
Minnesota Twins	87	Minnesota Twins	158	Pittsburgh Pirates	178
Toronto Blue Jays	90	Oakland Athletics	160	Tampa Bay Rays	167
		Tampa Bay Rays	160		
League average	194	League average	194	League average	227

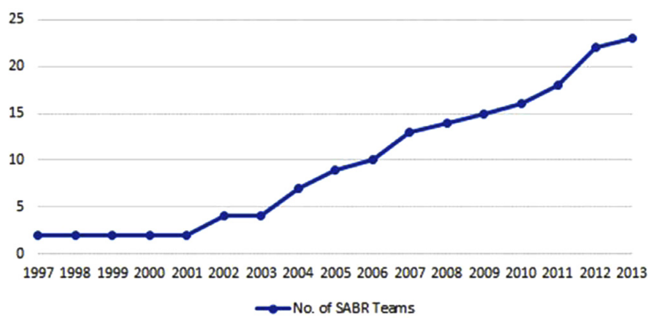

Fig. 2. No. of SABR teams by year.

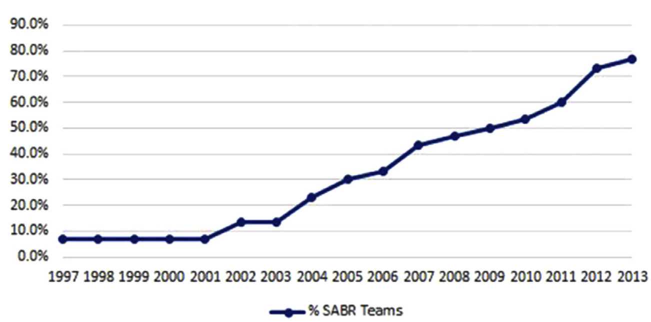

Fig. 3. %SABR teams by year.

This trend is captured in both Figs. 2 and 3. Table 1 and Fig. 1 indicate that between 1997 and 2001 there were only two Moneyball team (the Oakland Athletics and the San Diego Padres). In 2002, the last year in period 1, three MLB teams joined the data analytics revolution (the Boston Red Sox, the San Francisco Giants, and the Toronto Blue Jays) and one left until 2005 (the San Diego Padres). This was followed by an ever growing number of MLB teams who created baseball analytics departments over time, culminating in 23 teams (76.7% of MLB teams) with baseball analytics departments in 2013, as Figs. 2 and 3 illustrate.

One potential concern is whether the "Moneyball" effect disguises in reality a small market effect, essentially the result of the significant revenue disparity among major league baseball teams (which eventually led to the MLB instituting a revenue sharing program since 1996 [19]). Consequently, to control for this, small market teams are defined as those in the lowest sextile of total MLB team revenues, as disclosed by Forbes [21–23]. Table 3 provides the small market teams in each period. Comparing Tables 2 and 3 shows that the only small market "Moneyball" team in the first period, the Toronto Blue Jays, established its analytics department in 2002. In the second period, there are four small market "Moneyball" teams (the Oakland Athletics for the entire period, the Tampa Bay Rays since 2004, the Kansas City Royals since 2006, and the Pittsburgh Pirates in 2008). Table 3, combined with Table 2, shows that all small market teams in the third period are "Moneyball" teams (however, not all Moneyball teams are small market teams).

To examine individual GM performance, six MLB executives who rely on data analytics, and their teams over time, are tracked. These six GMs were identified following an internet search for all baseball executives in the period of the study and whether or not they are reported to rely on Sabermetrics. For simplicity, the term GM is used here to capture the executive who is in charge of the front office, whether it is a GM, the Manager of Baseball Research & Analytics, or the President of Baseball Operations. The data for 1994 and 1995 is dropped due to the MLB strike. In order to see whether there was a dramatic shift in 1997 (the year that Billy Beane reportedly started to use data analytics) data about GMs and their teams is collected from 1985 to 2013. This leads to a larger number of observations for these tests than the teams-oriented tests. Note that each of these GMs is specifically identified, and so the data captures whether the executive moved from one team to another over time. For example, Theo Epstein was the GM of the Boston Red Sox from 2003 until 2011, helping the team win the 2004 World Series for the first time since 1918, and then again in 2007. After leaving the Red Sox, Epstein has served as the president of the Chicago Cubs, and helped build the team that won the 2016 World Series, thus, breaking a 107-year draught (the longest in MLB history). The executives that were identified are provided in Table 4.

5.2. Descriptive statistics

Table 5 shows the descriptive statistics. Panel A provides descriptive statistics for all MLB teams, Panel B for small market teams, and Panel C for "Moneyball" teams. Panel A demonstrates that the overall mean *Win%* in the MLB, as can be expected, is around 50%. The *Win%* is about 50.5% for "Moneyball" teams, compared to 47% for small market teams. Also, these panels show that the mean *NormPay* over time is about 1 for all MLB teams, compared with 0.64 for small market teams and 0.97 for "Moneyball" teams, showing that "Moneyball" teams are slightly thriftier with their payroll than the rest of MLB and substantially less thrifty than smaller-markets teams. The overall team WAR, consistent with having a data analytics orientation, is higher (19.7) for "Moneyball" teams than for small market teams (15.73), and the entire MLB (18.99). The *Pay-For-Win* ($) (actual payroll in $ divided by the number of wins) is about $1.01 m for "Moneyball" teams, compared with $0.61 m for small market teams, and $0.91 m for the entire MLB.

Panel D shows the inflation in raw (before normalizing) payroll over time for the entire MLB: starting from an average team payroll of $53.7 in period 1, moving to $77.1 m in period 2, and to $94.3 m in period 3. Panel F shows the raw payroll over time for "Moneyball" teams: starting from an average team payroll of $58.2 m in period 1, moving to $77.8 m in period 2, and to $91.8 m in period 3. The reason that "Moneyball" teams' mean payroll is lower in period 1, and about the same in period 2, as the rest of MLB is that

Table 4
Data analytics GMs.

MB GM Number	Name	Teams
1	Andrew Friedman	Tampa Bay Rays, Los Angeles Dodgers
2	Billy Beane	Oakland Athletics
3	Keith Woolner	Cleveland Indians
4	Kevin Towers	San Diego Padres, Arizona Diamondbacks
5	Sandy Alderson	Oakland Athletics, San Diego Padres, New York Mets
6	Theo Epstein	Boston Red Sox, Chicago Cubs

Table 5
Summary statistics.

A. All MLB teams					
Variable	Obs	Mean	Std. Dev.	Min	Max
Win%	508	0.4998842	0.0718765	0.2654321	0.7160494
NormPay	508	1.005871	0.4225983	0.1719898	3.050332
TeamWAR	508	18.99941	7.914724	−4.5	44.4
Pay-For-Win ($)	508	910,140	415,009	136,350	2,729,163
B. Small-market teams					
Variable	Obs	Mean	Std. Dev.	Min	Max
Win%	91	0.4730252	0.0692353	0.3333333	0.5987654
NormPay	91	0.6357548	0.2239926	0.1844431	1.2189
TeamWAR	91	15.73077	7.413122	−4.5	33.5
Pay-For-Win ($)	91	613,556	225,903	163,715	1,255,064
C. "Moneyball" teams					
Variable	Obs	Mean	Std. Dev.	Min	Max
Win%	165	0.5045835	733174	0.3148148	0.6481481
NormPay	165	0.9718112	0.489855	0.1719898	3.050332
TeamWAR	165	19.7	7.819839	1	37.5
Pay-For-Win ($)	165	1,011,636	454,076	280,826	2,729,163
D. Mean raw payroll over time for the entire MLB($m)					
Variable	Obs	Mean	Std. Dev.	Min	Max
All years	508	73.90	36.20	10.60	232.00
1997–2002	178	53.70	22.70	10.60	126.00
2003–2008	180	77.10	33.60	14.70	208.00
2009–2013	150	94.30	39.70	17.90	232.00
E. Mean raw payroll over time for small market teams ($m)					
Period	Obs	Mean	Std. Dev.	Min	Max
All years	91	46.91	17.66	10.64	81.58
1997–2002	30	35.80	15.59	10.64	76.90
2003–2008	36	46.37	15.09	14.67	79.37
2009–2013	25	61.03	13.45	34.94	81.58
F. Mean raw payroll over time for SABR teams					
	Obs	Mean	Std. Dev.	Min	Max
All years	165	83.20	41.57	17.89	231.98
1997–2002	14	47.65	24.81	21.30	108.37
2003–2008	57	81.61	41.77	24.12	208.31
2009–2013	94	89.47	40.99	17.89	231.98

G. "Moneyball" teams vs. non-Moneyball teams over time

Variable	Period 1				Period 2				Period 3			
	*SABR*1		Non-*SABR*1		*SABR*2		Non-*SABR*2		*SABR*3		Non-*SABR*3	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Win%	52.5%	51.2%	49.8%	48.9%	50.9%	51.6%	49.6%	50.6%	49.9%	50.0%	50.2%	50.0%
NormPay	0.841	0.758	1.019	0.999	1.047	0.864	0.988	0.993	0.946	0.826	1.108	1.081
TeamWAR	19.42	18.30	18.97	18.80	19.68	18.20	18.68	18.80	19.75	19.30	17.73	17.95
Pay-For-Win ($)	560,128	484,903	669,916	643,505	973,858	925,792	932,195	907,833	1,101,790	1,015,256	1,266,162	1,207,151

Table 6
Pearson product–moment correlation table.

	Win%	*NormPay*	*TeamWAR*	*PayForWin ($)*
Win%	1.0000			
NormPay	0.3709	1.0000		
TeamWAR	0.7351	0.2937	1.0000	
Pay-For-Win ($)	0.0173	0.5193	0.0283	1.0000

Table 7
Spearman's rank correlation table.

	Win%	*NormPay*	*TeamWAR*	*PayForWin ($)*
Win%	1.0000			
NormPay	0.3580	1.0000		
TeamWAR	0.7279	0.2933	1.0000	
Pay-For-Win ($)	0.0142	0.3694	0.0203	1.0000

in 1997–2001 the only "Moneyball" teams are the Oakland Athletics and the San Diego Padres, whose budgets are relatively small, while in period 2 big spenders, such as the Boston Red Sox and the New York Yankees, become "Moneyball" teams and influence the number for the second period. In period 3, as 76.7% of the MLB is into "Moneyball", the mean payroll of "Moneyball" teams is below the MLB average. As expected, small market teams show substantially lower mean payroll in all three periods (Panel E).

Panel G shows that in the first 2 periods "Moneyball" teams have higher mean and median *Win%* than non-"Moneyball" teams. In the third period, 76.7% of the MLB uses baseball data analytics and, hence, the *Win%* of "Moneyball" and non-"Moneyball" teams converge. The table also shows that the median *NormPay* is lower for "Moneyball" teams than non-"Moneyball" teams in all three periods. The mean *NormPay* is lower for "Moneyball" teams than non-"Moneyball" teams in all period 1 and 3 but is higher in period 2, probably due to the addition of big spenders, such as the Boston Red Sox and the New York Yankees, to "Moneyball" teams in period 2, pushing the average up. *TeamWAR* on average is higher for "Monyball" than non-"Moneyball" teams in all three periods. Median *TeamWAR* is higher for "Moneyball" than non-"Moneyball" teams in period 1 and 3 but is lower in period 2 (18.2 vs. 18.8). Median and mean *Pay-For-Win ($)* is lower in periods 1 and 3 for "Moneyball" than non-"Moneyball" teams, and higher for period 2, probably due to the addition of big spenders, such as the Boston Red Sox and the New York Yankees, to "Moneyball" teams in period 2

5.3. Correlations

Table 6 provides the Pearson Product-Moment correlations and shows that the *Win%*'s is most correlated with *TeamWAR* (73.51%), and the least with *Pay-For-Win* (1.73%). Interestingly, the correlation of *Win%* with *NormPay* is 37.09%, about half of its correlation with *TeamWAR* (73.51%). This implies that while MLB team payrolls affect winning their pursuit of *TeamWAR* is of utmost importance. Moreover, the correlation between *NormPay* and *TeamWAR* is only 29.37%, possibly indicating inefficiencies in the way that MLB teams create rosters. *Pay-For-Win*'s correlation with *NormPay*'s is 51.93%, while its correlation with *TeamWAR* is only 2.83%, possibly indicating perhaps again MLB payroll inefficiencies. Similar results are obtained for the Spearman's Rank Correlations, shown in Table 7.

5.4. Billy Beane's performance over time

As discussed in Section 4.1, Billy Beane's strategy is often measured by the *Pay-For-Win*. Thus, I plot Beane's Win % against normalized payroll, as shown in Fig. 3:

Table 8
Tests of model 1.

VARIABLES	All Years (5) Win
SNP	0.0180***
	(3.340)
TeamWAR	0.00638***
	(28.40)
SABR1	0.0925***
	(3.123)
*SABR1*SNP*	−0.0411
	(−0.974)
SABR2	0.0211
	(0.920)
*SABR2*SNP*	0.0500**
	(2.351)
SABR3	0.0298
	(1.272)
*SABR3*SNP*	0.0180
	(0.776)
SM1	−0.0230
	(−1.436)
*SM1*SNP*	0.00332
	(0.0857)
SM2	−0.0358*
	(−1.835)
*SM2*SNP*	0.127**
	(2.208)
SM3	0.0233
	(0.820)
*SM3*SNP*	−0.0381
	(−0.482)
Constant	0.321***
	(17.99)
Observations	800
Non-SABR Teams dummies	Yes
R-squared	0.616

Robust *t*-statistics in parentheses.
*** $p < 0.01$,
** $p < 0.05$,
* $p < 0.1$.

Fig. 4 shows that Billy Beane generates, consistent with Hypothesis 3, a higher Win% relative to normalized payroll, prior to 2003 (the cluster within the red oval) than afterwards. Moreover, it shows a vertical upwards trending trajectory of Win % at the same normalized payroll. In contrast, after 2003 (the cluster within the yellow circle), consistent with the idea that the advantage provided by such organizational knowledge is comparative, as Hypothesis 4 predicts, his performance is sporadic.

6. Results and discussion

The first rule about fight club is you don't talk about fight club… The second rule about fight club is you don't talk about fight club [45].

6.1. Results and discussion of Hypotheses 1 and 2

The results of Model 1, shown in Table 8, demonstrate that the scaled payroll, *SNP*, has a positive and highly significant effect on *Win%*. The results also demonstrate that *TeamWAR* has a positive and highly significant effect on *Win%*, illustrating the importance of WAR within baseball analytics.

Consistent with Hypothesis 1, Table 8 shows an advantage for "Moneyball" teams in periods 1 and 2, shown by the highly significant positive coefficient of the fixed effect in period 1, *SABR*1, and the significant positive coefficient for the interaction variable in period 2, *SNP*SABR*2. Consistent with Hypothesis 2, these effects disappear altogether in period 3.

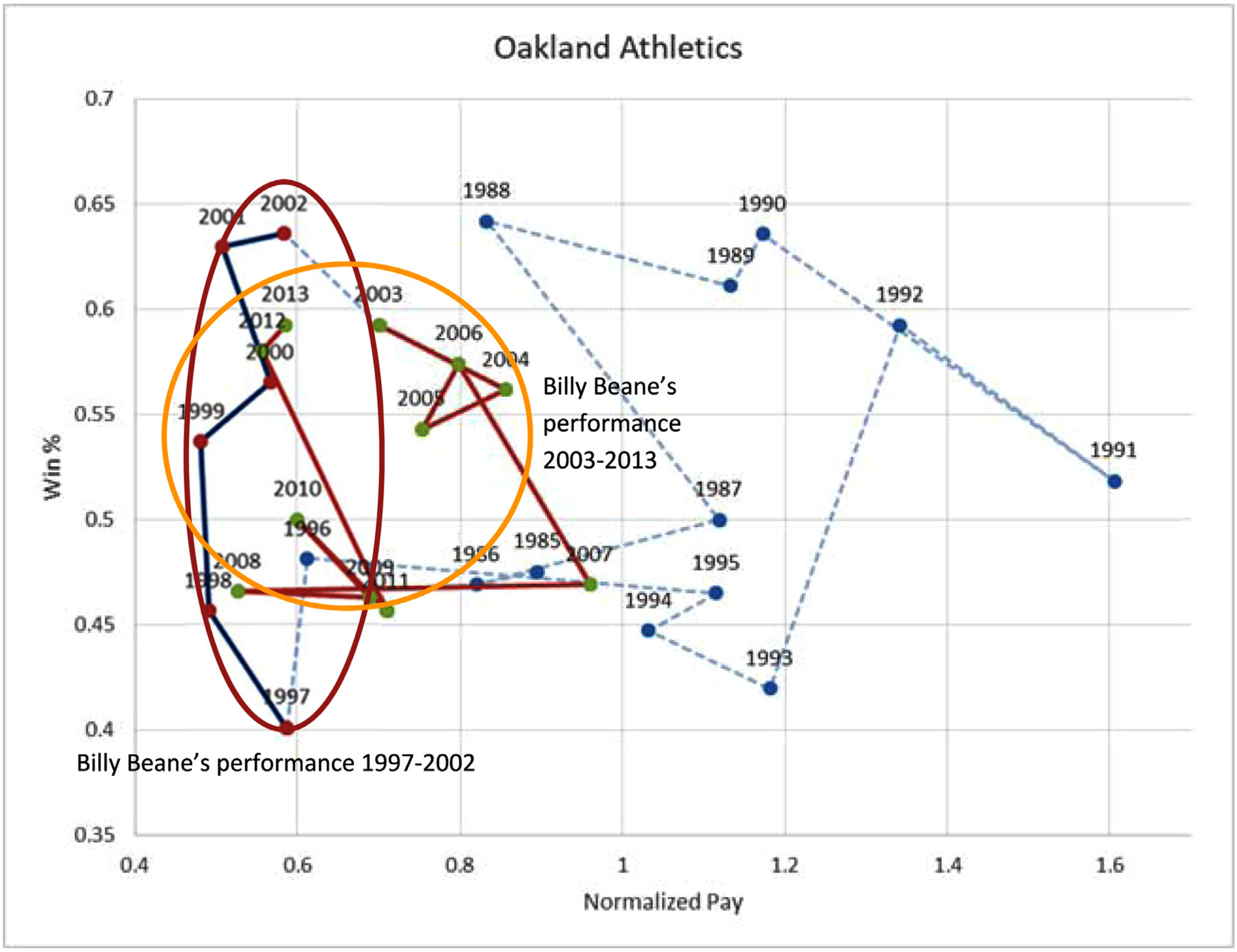

Fig. 4. Billy Bean's performance pre-"Moneyball" versus post-"Moneyball".

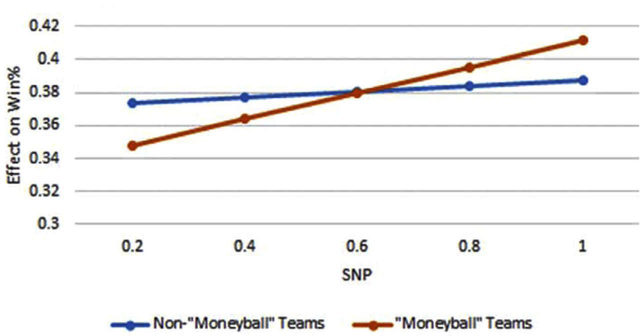

Fig. 5. Moneyball teams effects period 2.

Table 9
"Moneyball" teams versus the rest of MLB by period (model 2).

Variables	1997–2002 (2a) Win%	2003–2008 (2b) Win%	2009–2013 (2c) Win%
NormPay	0.0188	0.0201	0.0250
	(0.857)	(0.832)	(1.552)
TeamWAR	0.00579***	0.00691***	0.00661***
	(10.64)	(13.10)	(14.58)
$SABR_t$	0.0975***	0.0428*	−0.00428
	(2.656)	(1.935)	(−0.182)
$NormPay*SABR_t$	−0.0208	0.0137	−0.0101
	(−0.489)	(0.526)	(−0.516)
SM_t	−0.0597	−0.0585*	−0.000991
	(−0.879)	(−1.895)	(−0.0209)
$NormPay*SM_t$	0.0143	0.0965**	0.0125
	(0.251)	(2.039)	(0.174)
Constant	0.320***	0.294***	0.357***
	(21.38)	(11.20)	(17.51)
Observations	178	180	150
Non-SABR Teams dummies	Yes	Yes	Yes
R-squared	0.739	0.703	0.596

Robust *t*-statistics in parentheses.
*** $p < 0.01$,
** $p < 0.05$,
* $p < 0.1$.

Table 8 demonstrates that small market teams effects are mixed and occur only in period 2, shown by a weakly significant negative fixed effect (*SM*2) and a significant positive interaction effect *SM2***SNP*. To examine this in greater depth, a simulation of the overall net effect of *SNP* on *Win%* for small market teams in period 2 was conducted, as depicted in Fig. 5. The Figure shows that small market teams have an advantage over other teams only when their SNP exceeds 0.2. Examining the data shows, however, that 55% the small market teams observations in period 2 (20 observations out of 36) have SNP = 0.2, and another 33% (12 observation) are in the 0.4 SNP range, where the advantage of small market teams is very small.

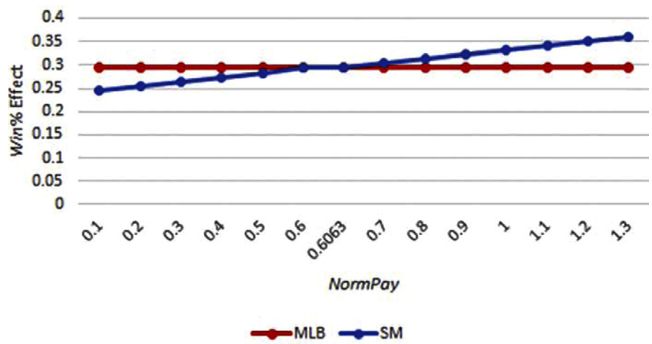

Fig. 6. Small market team effects in period 2.

The results of Model 2, shown in Table 9, demonstrate that in each of the three periods there is a positive relationship between *NormPay* and *Win%*. This coefficient however is not significant except for small market teams in period 2, shown by the coefficient of SNP^*SM_2, perhaps due to their necessity to be efficient and effective with their payroll.

Moreover, Models 2a–2c show, as expected, a positive and highly significant relationship between the control variable *WAR* and *Win%* in each sub-period (Models 2a–2c). This, consistent with Model 1, demonstrates the importance of *WAR* for baseball analytics.

The table shows, consistent with Hypothesis 1, a highly significant positive coefficient for the fixed "Moneyball" effects, $SABR_1$, in period 1 (Model 2a), a weakly significant positive coefficient for period 2, $SABR_2$ (Model 2b), and a negative non-significant coefficient for period 3, $SABR_{32}$ (Model 2c). Moreover, consistent with Hypothesis 2, the fixed effect coefficient for "Moneyball" teams is more than double in period 1 (0.0975) than period 2 (0.0428) and then becomes negative (but non-significant) in period 3 (−0.00428). The interaction variable, $NormPay^*SABR_t$, however, is not significant for any of the three periods, demonstrating that while "Moneyball" teams have had a higher intercept for the regression than other MLB teams, the slope is the same.

Table 9 also shows, consistent with Model 1, that small team effects, as revealed by SM_t and $NormPay^*SM_t$, are mixed and occur only in period 2, where the fixed effect coefficient, SM_2, is weakly significant and negative, and the interaction effect coefficient, $NormPay^*SM_2$, is significantly positive. Fig. 6 depicts the numerical analysis of these coefficients, which shows that small market teams outperform other teams only when *NormPay* exceeds 0.6063. Analysis of the data shows that there are 36 observations with small market teams, 19 of them fall below 0.6063 and 17 above it. As such, it implies the small market effect in period 2 are mixed.

In summary, Tables 8 and 9 demonstrate, consistent with Hypotheses 1 and 2, that "Moneyball" teams have had a competitive pay-performance strategic advantage when data analytics related organizational knowledge was proprietary but the advantage has gradually vanished once the link between data-analytics teams' knowledge and their sustained competitive advantage became unambiguous. Moreover, these results demonstrate that small market effects wash out.

6.2. Results and discussion of Hypotheses 3 and 4

Table 10 shows the results for models 3 and 4, which test for individual "Moneyball" (i.e., those GMs identified in Table 4) GMs effects. Model 3 looks at the entire period without partitioning into the pre- and post-"Moneyball" periods. It shows that there is a positive and significant relationship between Win % and normalized payroll in the league (*NormPay*), and for Andrew Friedman (*GM*1 and *GM*1**NormPay*). Billy Beane (*GM*2), interestingly, has a positive and significant fixed effect dummy (*GM*2) and a significant negative interaction effect (*GM*2**NormPay*). As such, this shows that while Billy Beane had a better fixed effect, essentially a 0.547 constant vs. 0.429 for MLB, the use of data analytics has resulted in a negative variable effect (−0.0429 for Billy Beane vs. 0.0681 for the rest of MLB). However, Oakland's *NormPay* was low (moreover, the A's had the lowest *Pay-For-Win* ($) in MLB during that period) and, thus, the fixed effect was more than enough to offset the negative slope. Beane's mentor Sandy Alderson (*GM*5) had, however, subpar performance relative to the league as his constant effect is not significantly different than MLB while the variable effect is significantly below the league's (0.0371 versus MLB's 0.0681). Keith Woolner's (*GM*3) results are very similar to Billy Beane, significantly positive fixed effect and significantly negative variable effect. Theo Epstein (*GM*6) has a weakly significant variable effect but no fixed effect. Kevin Towers however shows no difference whatsoever from the rest of the league.

Model 4 adds to model 3 by partitioning the data into the pre-"Moneyball" vs. post-"Moneyball" sub-periods. The results show

Table 10
Model 3 and 4 (individual GM effects).

Variables	(3) Win	(4) Win
NormPay	0.0681***	0.0725***
	(8.459)	(5.989)
MBYears		0.00592
		(0.364)
*MBYears*NormPay*		−0.00841
		(−0.570)
GM1	−0.0721***	−0.0721***
	(−8.560)	(−5.903)
GM1 NormPay*	0.257***	0.261***
	(31.97)	(25.19)
GM2	0.118***	0.111***
	(12.79)	(8.943)
GM2 NormPay*	−0.111***	−0.0968***
	(−13.83)	(−7.992)
MBGM2		−0.00625
		(−0.385)
MBGM2 NormPay*		−0.000177
		(−0.0120)
GM3	0.154***	0.152***
	(18.44)	(12.36)
GM3 NormPay*	−0.197***	−0.193***
	(−24.41)	(−18.56)
GM4	0.00275	−0.00836
	(0.0363)	(−0.686)
GM4 NormPay*	−0.00960	−0.0227*
	(−0.113)	(−1.875)
MBGM4		0.00256
		(0.0147)
MBGM4 NormPay*		0.0408
		(0.178)
GM5	0.0196	−0.0105
	(1.180)	(−0.842)
GM5 NormPay*	−0.0310**	−0.00981
	(−2.419)	(−0.810)
MBGM5		0.0462***
		(2.799)
MBGM5 NormPay*		−0.0394**
		(−2.516)
GM6	−0.153	−0.154
	(−1.624)	(-1.635)
GM6 NormPay*	0.105*	0.109**
	(1.893)	(1.980)
Constant	0.429***	0.426***
	(45.44)	(33.47)
Observations	769	769
Team Fixed Effects	Yes	Yes
R-squared	0.1667	0.1676
Number of TeamNo	30	30

that, similar to model 3, there is a positive and significant relationship between *Win %* and normalized payroll in the league (*NormPay*), and that Andrew Friedman's has significant fixed and interaction dummies (*GM*1 and *GM*1**NormPay*, respectively). The table also shows that Billy Beane (*GM*2) has significantly positive fixed and significantly negative variable effects only before 2002. Keith Woolner (*GM*3), who became a GM only after 2007, has significant positive fixed and significant negative variable effects. Sandy Alderson (*GM*5) has no advantage compared to the league prior to 2002 but after 2002 has a significantly better constant (*MBGM*5) and a significantly worse variable effect (*MBGM*5**NormPay*) than the league. Theo Epstein (*GM*6), who assumed his position only in 2003 (the year that the book was published), shows a significant advantage in his variable effect but no significant fixed effect. In essence, the analysis supports Hypothesis 3 and shows that Billy Beane and his mentor, Sandy Alderson, fared better than the rest of MLB before the publication of the book in 2002. Thus, showing the value of the data analytics related organizational knowledge in the competitive MLB marketplace. This analysis also provides support for Hypothesis 4 by showing that once such information becomes public its value diminishes. In addition, some GMs who started only after 2002 (i.e., Andrew Friedman, Keith Woolner and Theo Epstein) demonstrate the value of their proprietary information by having significant advantages over MLB.

6.3. Robustness tests

First, raw payroll is used instead of *NormPay* as one of the independent variables, yielding very similar results to the ones above.

Second, "Moneyball" teams are identified using a quantitative test, instead of the year in which the baseball analytics department was established for each MLB team. The quantitative criterion used is that "Moneyball" teams are those located in the top third of the league in winning and the lowest third of *Pay-For-Win*. The idea behind this robustness test is that "Moneyball" teams maximize wins while minimizing the cost per win [24]. The results of this test are consistent with Hypothesis 1 and 2. It should be noted that while there is some intersection between the main test and the robustness test in terms of "Moneyball" team identification, it is not complete as the identification procedure in the robustness test is cruder. The teams that show up as "Moneyball" teams in both procedures are: the Oakland Athletics (in all three periods under both sets of tests), the San Francisco Giants (period 1 in the robustness test vs. 2002 and beyond in the main test), the Cleveland Indians (period 2 in the robustness test vs. 2007 and beyond in the main test), Tampa Bay Rays (period 3 of the robustness test vs. 2004 and beyond in the main test), Cincinnati Reds (period 3 in both sets of tests), Texas Rangers (period 3 of the robustness tests vs. 2011 and thereon in the main test). Moreover, under this robustness test an earlier period is tested (1990–1996, excluding the strike year). The results of the tests demonstrate, as expected, no pay-performance advantages to teams identified in this pre-"Moneyball" period.

Third, to determine whether the predictions of the models all stem from *TeamWAR* (the most highly correlated variable with *Win%*), the models are run with *TeamWAR* as the only explanatory variable (Table 11) and then with all explanatory variables except *TeamWAR* (Table 12). Tables 11 and 12 show that, while *TeamWAR* is an important explanatory variable, the other variables are important too. This is revealed by the fact that the other explanatory variables add to the statistical power of the regression (the R^2 in Table 11 is lower than in Tables 8 and 9, and the R^2 in Table 12 is meaningful). Furthermore, Table 12 shows the same signs of the coefficients of the explanatory variables as Table 9. Out of the all "Moneyball" variables only $SABR_1$ has a significant coefficient in Tale 12, compared to Table 9 where $SABR_1$ has a highly significant coefficient, and $SABR_2$ has a weakly significant positive coefficient.

Table 11
Regressions with *TeamWAR* as the only explanatory variable.

Variables	All Years (1) Win%	1997–2002 (2a) Win%	2003–2008 (2b) Win%	2009–2013 (2c) Win%
TeamWAR	0.00661***	0.00686***	0.00649***	0.00675***
	(31.99)	(16.18)	(13.49)	(15.46)
Constant	0.375***	0.369***	0.377***	0.372***
	(86.65)	(41.65)	(35.49)	(40.77)
Observations	800	178	180	150
R-squared	0.540	0.596	0.470	0.571

Robust *t*-statistics in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 12
Regressions all explanatory variable except *TeamWAR*.

Variables	1997-2002 (2a) Win%	2003-2008 (2b) Win%	2009-2013 (2c) Win%
NormPay	0.0336	0.0241	0.0424*
	(1.109)	(0.644)	(1.868)
$SABR_t$	0.118**	0.0242	−0.0253
	(2.037)	(0.771)	(−0.741)
$NormPay*SABR_t$	−0.00428	0.0411	0.0276
	(-0.0726)	(1.046)	(1.008)
SM_t	−0.0447	−0.0580	0.0465
	(−0.501)	(−1.385)	(0.558)
$NormPay*SM_t$	−0.0107	0.0820	−0.0529
	(−0.141)	(1.458)	(−0.426)
Constant	0.387***	0.418***	0.455***
	(21.05)	(12.00)	(16.27)
Observations	178	180	150
R-squared	0.514	0.388	0.129
Non-SABR Teams dummies	Yes	Yes	Yes

Robust *t*-statistics in parentheses.
*** $p < 0.01$,
** $p < 0.05$,
* $p < 0.1$.

In addition, small market effects are not demonstrated at all in Table 12, compared to Table 9 where period 2 shows such effects. Last, the coefficient of *NormPay* is weakly significant in period 3 in Table 12, compared to Table 9 where it is not significant in any period . Overall, the results in Tables 11 and 12 are consistent with Hypotheses 1 and 2.

7. Concluding remarks

The broad question that this paper studies is whether data analytics organizational knowledge provides a strategic advantage in a competitive marketplace, and whether this advantage, if any, is absolute or comparative. The advantage from data analytics is tested by analyzing whether "Moneyball" teams and GMs enjoy a superior pay-performance relative to their competition. To test whether the advantage from data analytics is absolute or comparative I analyze whether it has persisted over time, or vanished once it became publicly available.

The first hypothesis that I examine is whether data analytics has provided a pay-performance advantage to "Moneyball" teams over their competition. To test whether this advantage is comparative or absolute (the second Hypothesis), I examine whether it has persisted after this information has become publicly available. The results show that "Moneyball" teams have had an advantage in the period before their use of data analytics was "outed" (1997–2002), this advantage is weakened in the period following the outing of this knowledge (2003–2008) and dissipates afterwards (2009–2013). Therefore, showing the strategic value of such proprietary information in the competitive market place and that the value of this information is comparative, rather than absolute.

Next, individual GMs are analyzed. This analysis shows that Billy Beane and his mentor, Sandy Alderson, fare better before the publication of the book in 2002, thus, demonstrating the value of this proprietary organizational knowledge. This pay-performance advantage vanishes once the information has become publicly available, thus, supporting the idea that the strategic advantage derived from data analytics related organizational knowledge is comparative and not absolute. The results also show that some GMs hired only in 2002 and afterwards (Andrew Friedman, Keith Woolner and Theo Epstein) have significant advantages over the rest of MLB but, because of the time they were hired, a comparison of the pre-"Moneyball" and post-"Moneyball" periods cannot be conducted for them.

The study, as with any research, suffers from several limitations that could be addressed in an extension. First, the identification of "Moneyball" teams is based on a structured search of the year that each MLB team has established a baseball data analytics department. This search however could potentially miss some salient information on the establishment of those departments. Consequently, identifying "Moneyball" teams can be further extended by examining the types of trades that teams make (for example, prospects versus established players), the types and quantity of free agents that teams sign, or the fraction of the roster that is "homegrown".

Second, the concept of "Moneyball" executives that is used here is based on an internet search of their reputation and therefore lacks objectivity. A possible executive-focused extension to this study could apply the ideas of Hambrick and Mason [32], who state that organizational outcomes, such as strategic choices and performance, are a reflection of the backgrounds of upper management. Accordingly, such an extension should carefully look at the background of front office executives, and changes in front offices over time.

Acknowledgements

The author thanks the insights and help from Opher Baron, Jae Chang, Michael Marin, Richard Mills, Partha Mohanram, Christopher Small, Dushyantkumar Vyas and insightful comments from Jeffrey Callen, Alex Edwards, Tom Galimanas, Arieh Gavious, David Lutterman, Matt Mitchell, Peri Mottath, Hilla Fogel Yaari, Morgan Pampe, Marshall Vance, Kent Womack, and Paul Zarowin, as well as participants in the Rotman Accounting Workshop and the Management Workshop in Ben Gurion University. The author also gratefully acknowledges the financial support from the Edward J. Kernaghan Professorship in Financial Analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.omega.2018.11.010.

Appendix 1 – WAR and Its Calculation

WAR is widely viewed as the most important statistic in baseball [10]. The idea behind it is to measure the number of incremental number of wins of a team that a player provides relative to the expected number of wins by a replacement level player. Replacement players are baseline players, i.e., those who are freely available and can be replaced at the league's minimum salary (normally through waivers, or from the minor leagues). WAR is available from FanGraphs (fWAR, [20]) and Baseball-Reference (bWAR or rWAR, [4]), or Baseball Prospectus (WARP). The three sources calculate replacement level the same way but are slightly different, for example, in the way that they calculate the WAR for pitchers (for a more detailed explanation of those difference please see the WAR Comparison Chart in Baseball-Reference, n.d.). The paper uses the calculations of WAR from FanGraphs, i.e., fWAR. All three sources of WAR calculate an aggregate team WAR. As previously mentioned, the study uses the fWAR calculations.

For positional players FanGraphs first calculates the runs they produced above the league's average through batting, base running, and denying through defense runs for opponent teams (Fielding Runs). This number is then adjusted for the position (e.g., catcher vs. first baseman), and the league (American vs. National). Next, the sum of runs is scaled based on comparison to a replacement player (Replacement Runs). Lastly, FanGraphs takes the adjusted sum and divides it by the runs that led on average to a win in that season. As such, WAR for position players is calculated by FanGraphs as

$$\frac{\text{Batting Runs} + \text{Base Running Runs} + \text{Fielding Runs} + \text{Positional Adjustment} + \text{League Adjustment} + \text{Replacement Runs}}{\text{Runs Per Win}}$$

For pitchers WAR is a more complicated calculation that uses Fielding Independent Pitching (FIP), which estimates the number of runs that a pitcher prevented, independent of the team defence. Pitchers' WAR is adjusted for the ball park where the game was played, and the number of innings that the pitcher threw.

Bopp [8] reports that WAR's for 38% of MLB players are between 0 and 1, 35% between -1 and 0, and only 19% are between 1 and 4.

Several Sabermetric studies demonstrate that team WAR (i.e., the cumulative WAR of all players in a baseball team) is highly predictive of the team's wins [9,17,33,50].

Appendix 2 –Variable Definition

GM_iDummy - fixed effect variable (intercept dummy) for *GMi* (the list is provided in Table 4)

MBGM – fixed effect variable (intercept dummy) identifying all 7 "Moneyball" GMs as a group

$MBGM_i$ – fixed effect variable (intercept dummy) denoting fixed effect of GM_i after 2002

MBYears - fixed effect variable (intercept dummy) denoting the era after 2002

$NormPay_j$ - $\frac{Team\ j\ Payroll}{Average\ Payroll\ in\ MLB\ baseball\ that\ year\ without\ team\ j}$

Pay-For-Win ($) - actual payroll in $ divided by the number of wins

$SABR_t$ – fixed effect variable (intercept dummy) denoting a "Moneyball" team in period t (t = 1,…,3).

SM_t – fixed effect variable (intercept dummy) for smaller market teams in period t (t = 1, …,3).

SNP – Scaled *NormPay*. Using quintile ranks, *NormPay* is scaled to be between zero and one.

TeamWAR – the aggregate WAR of a team in a given year

WAR – Wins Above Replacements, discussed in length in Appendix 1 above

Win% – the percentage of games won by a team in a given season.

References

[1] Abbott H2010. The Simpsons go sabermetric ESPN.go.com. Available on http://espn.go.com/blog/truehoop/post/_/id/20319/the-simpsons-go-sabermetric.
[2] Barney J. Firm resources and sustained competitive advantage. J Manag 1991;17(1):99–120.
[3] Baseball-Reference.com. Team wins. Available on http://www.baseball-reference.com/leagues/MLB/.
[4] Baseball-Reference.com. WAR comparison chart.n.d. Available on http://www.baseball-reference.com/about/war_explained_comparison.shtml.

[5] Baumer B2015. In a Moneyball world, a number of teams remain slow to buy into sabermetrics. Special to ESPN. February 23, 2015. Available on http://www.espn.com/espn/feature/story/_/id/12331388/the-great-analytics-rankings.
[6] Bertsimas D, Bradlow E, Gans N, Gupta A. Introduction to the special issue on business analytics. Manage Sci 2014;60(6):1351.
[7] Bickel JE. On the decision to take a pitch. Decis Anal 2009;6(3):186–93 September.
[8] Bopp J. How common is a one-win above replacement player? http://www.beyondtheboxscore.com/2010/11/16/1813082/how-common-is-a-one-win-above-replacement-player.
[9] Cameron D2008. Win values explained: part three http://www.fangraphs.com/blogs/explaining-win-values-part-three/.
[10] Catania J2013. Making the Case for WAR as Baseball's Most Perfect Statistic https://bleacherreport.com/articles/1642919-making-the-case-for-war-as-baseballs-most-perfect-statistic.
[11] Chan TCY, Fearing D. Process flexibility in baseball: the value of positional flexibility. accepted to. Manage Sci 2017 2017.
[12] Chettiar IM. More police, managed more effectively, really can reduce crime. The Atlantic. 2015. February 11, 2015. Available on https://www.theatlantic.com/national/archive/2015/02/more-police-managed-more-effectively-really-can-reduce-crime/385390/ .
[13] Davenport TH, Harris JG. Competing on analytics: the new science of winning. Boston: Harvard Business School Press; 2007.
[14] Deli D. Assessing the relative importance of inputs to a production function: getting on base versus hitting for power. J Sports Econ 2013;14(2):203–17.
[15] Demmink H. Value of stealing bases in Major League Baseball: stealing runs and wins. Public Choice 2010;142(3–4):497–505.
[16] Doumpos M, Zopounidis. Special issue business analytics. Omega 2016;59(A):1–130.
[17] DuPaul G. 8 August 2012. What is WAR good for? Available on http://www.hardballtimes.com/what-is-war-good-for/.
[18] Theo Epstein. 2018. GM at Boston Red Sox Available on. (retrieved September 2, 2018) https://www.linkedin.com/in/theo-epstein-b6659a18/.
[19] FanGraphs.com Revenue Sharing. Available on https://www.fangraphs.com/library/business/revenue-sharing/.
[20] FanGraphs.com What is WAR?. Available on http://www.fangraphs.com/library/misc/war/.
[21] Forbes â€MLB Team Valuations 2002 Available on https://www.forbes.com/home/free_forbes/2002/0415/092tab2.html.
[22] Forbes. Special report: the business of baseball available on https://www.forbes.com/lists/2009/33/baseball-values-09_The-Business-Of-Baseball_Rank.html.
[23] Forbes-MLB Team Valuations 2013. Available on http://www.thesportsadvisorygroup.com/resource-library/business-of-sports/forbes-mlb-team-valuations-2013/.
[24] Foster, G. Oreilly, N. Lippert, R. Shimizu, C. and Udset, K. 2014. Billy Beane and the Oakland Atheltics (A): disruptive Innovation in Major League Baseball. Stanford Business Case SPM-53(A).
[25] Fry MJ, Ohlmann JW. Introduction to the special issue on analytics in sports, part I: general sports applications. Interfaces 2012;42(2):105–8.
[26] Fry MJ, Ohlmann JW. Introduction to the special issue on analytics in sports, part II: Sports scheduling applications. Interfaces 2012;42(3):229–31.
[27] Futterman M. Baseball after moneyball. Wall Street J 2011. Available on http://www.wsj.com/articles/SB10001424053111903791504576584691683234216.
[28] Gelles Sarah. 2018. Director, Analytics & Major League Contracts at Baltimore Orioles. (retrieved on September 2, 2018) https://www.linkedin.com/in/sarah-gelles-882a9a1a.
[29] Gold AH, Malhotra A, Segars AH. Knowledge management: an organizational capabilities perspective. J Manage Inf Syst 2001;18(1):185–214.
[30] Green J, Hand JRM, Soliman M. Going, going, gone? the apparent demise of the accruals anomaly. Manage Sci 2011;57(5):797–816 May.
[31] Hakes JK, Sauer RK. An economic evaluation of the Moneyball hypothesis. J Econ Perspect 2006;20(3):173–86 (Summer, 2006).
[32] Hambrick DC, Mason PA. Upper echelons: the organization as a reflection of its top managers. Acad Manage Rev 1984;9(2):106–93.
[33] Keller JJ. MLB: an update on the correlation between fWAR and wins. Statliners 2014. November 21, 2014. Available on http://statliners.com/2014/11/21/mlb-update-correlation-fwar-wins/.
[34] Sean Lahman's Baseball Database. Available on http://www.seanlahman.com/baseball-archive/statistics/.
[35] Lewis Michael. Moneyball: the art of winning an unfair game. New York: Norton; 2003.
[36] Liberatore M, Luo W. The analytics movement: implications for operations research. Interfaces 2010;40(4):313–24.
[37] Liebeskind J. Knowledge, strategy, and the theory of the firm. Strategic Manage J 1996;17:93–107.
[38] Lo AW. The adaptive market hypothesis: market efficiency from an evolutionary perspective. J Portfolio Manage 2004;5(30):15–29.
[39] Lo AW. Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. J Invest Consul 2005;7(2):21–44.
[40] Lo AW. Adaptive markets: financial evolution at the speed of thought, Princeton, NJ: Princeton University Press; 2017. 2017.
[41] Long CD. Chief data scientist at headlamp software https://www.linkedin.com/in/octonion/.
[42] Milgram A. Moneyballing criminal justice. The Atlantic. 2012. June 20, 2012. Available on https://www.theatlantic.com/national/archive/2012/06/moneyballing-criminal-justice/258703/.
[43] Miller G, Weatherwax M, Gardinier T, Abe N, Melville P, Pendus C, Jensen D, Reddy CK, Thomas V, Bennett J, Anderson G, Cooley B. Tax collections optimization for New York State. Interfaces 2012;42(1):74–84.
[44] Mohanram PS. Analysts' cash flow forecasts and the decline of the accruals anomaly. Contemp Account Res 2014;31(4):1143–70.
[45] Palahniuk C. Fight club. New York: Owl Books; 1999.
[46] Pinto K2015. USD Law Alumnus Theo Epstein '00 (JD) Featured in Chicago Tribune Article Available on http://www.sandiego.edu/news/law/detail.php?_focus=52873.
[47] Radhakrishnan S, Wu S. Analysts' cash flow forecasts and accrual mispricing. Contemp Account Res 2014;31(4):1191–219.
[48] Scully G. Pay and performance in Major League Baseball. Am Econ Rev 1974;64:915–30.
[49] Schwert GW. Anomalies and market efficiency. In: Constantinides G, Harris M, Stulz R, editors. Handbook of the economics of finance, 15. North-Holland; 2003. p. 937–72. Chapter.
[50] Sullivan J. Does projected team WAR actually mean anything. FanGraphs 2014. December 18, 2014. Available on http://www.fangraphs.com/blogs/does-projected-team-war-actually-mean-anything/.
[51] Thaler R. Advances in behavioral finance. ed. New York: Russell: Sage Foundation; 1993.
[52] Troilo M, Bouchet A, Urban TL, Sutton WA. Perception, reality and the adoption of business analytics: evidence from north american professional sports organizations. Omega 2017;59:72–83.
[53] Wikipedia. Theo Epstein 2018. Available on https://en.wikipedia.org/wiki/Theo_Epstein.