In Search of the Holy Grail: Team Chemistry and Where to Find It

Scott A. Brave* R. Andrew Butters† Kevin A. Roberts‡

June 30, 2017

Abstract

Measuring “chemistry” and its effect on team performance has proven to be an elusive concept in sports analytics. In order to shed further light on this topic, we apply advanced statistical techniques designed to capture player complementarities, or “network” relationships, between teammates on Major League Baseball teams. Using FanGraphs’ wins-above-replacement metric (fWAR) over the 1998-2016 seasons and a spatial factor model embodying the strength of teammate’s on-the-field interactions, we show that approximately 44 percent of the unexplained variation in team performance by fWAR can be explained by chemistry. By building a set of novel individual player metrics which control for a player’s effect on his teammates, we are then able to develop some “rules-of-thumb” for team chemistry that can be used to guide roster construction.

Keywords: Team Chemistry, Network Analysis, Wins-above-replacement, Spatial PCA

*Federal Reserve Bank of Chicago, 230 S. LaSalle St. Chicago, IL 60604. E-mail: sbrave@frbchi.org, phone: (312) 322-5784 (Corresponding Author).
†Business Economics and Public Policy, Kelley School of Business, Indiana University, 1309 E. Tenth St. Bloomington, IN 47405. E-mail: rabutter@indiana.edu.
‡Federal Reserve Bank of Chicago, 230 S. LaSalle St. Chicago, IL 60604. E-mail: keroberts231@gmail.com.
1 Introduction

Chemistry, intangibles, and a whole that is greater than the sum of its parts: These are the euphemisms that often get thrown around in the sports media in an effort to rationalize how a team made up of seemingly inferior players manages to outperform another that on paper looks unbeatable. While these analogies are plentiful, little consensus exists on the proper way to attribute a team’s performance to its chemistry. At the heart of the argument, however, is the notion that some teams just manage to get the “right fit” of players together, while other teams—potentially still strong (or even superior) on an individual basis—do not. What defines the “right fit” can often be very subjective. Here, we provide a formal definition based on player complementarities which allows us to objectively measure team chemistry in Major League Baseball.

Evaluating the degree to which a variety of inputs – in our case, different position players on a sports team – are complementary or substitutable in production (e.g. of team wins) is a topic that economists have wrestled with for just under a century (Hicks (1932) and Robinson (1933)). We appeal to this tradition and apply advanced statistical techniques designed to capture player complementarities, or “network” relationships between teammates, to measure the degree of team chemistry. Finding a large degree of complementarities across players on the same team provides scope for the hypothesis that team chemistry plays a fundamental role in team success in baseball. Similarly, finding individual players whose presence routinely complements their teammates allows for the identification of its sources.

We begin our analysis by using FanGraphs’ wins-above-replacement metric ($fWAR$) to construct player productivity residuals for the 1998-2016 seasons. These residuals reflect the difference between the expected and actual number of team wins that can be attributed to each player in a given season. When aggregated across teammates, they measure the difference between a team’s actual win count and its expected wins based solely on individual player performances. If $fWAR$ was a comprehensive measure of each players’ contribution to team wins as it is advertised and players were perfectly substitutable along this dimension, the residuals for each team would sum to zero. However, this is not the case, with 20 percent of the variation in wins across teams left unexplained according to our productivity residuals.

From this unexplained variation in the win-loss ledger of MLB teams, we then isolate the element of team wins arising from teammate interactions as opposed to potential mismeasurement in $fWAR$. To measure the strength of teammate interactions, we take into account several dimensions of teammates’ on-the-field relationships, weighting more heavily pairings: 1) that play more often (taking into account both past and present playing time), and 2) that are characterized by the network relationships that exist between hitters in a team’s lineup and defensive positions. For instance, the interactions of hitters who bat in adjacent positions at the top of the lineup are given more weight than the interactions of hitters who bat at the top
vs. the bottom of the lineup. Similarly, the pitcher-catcher defensive relationship is given more weight than any other defensive pairing on the field.

Measuring teammate interactions in this way lends itself to the use of a spatial factor model to decompose our player productivity residuals into two separate unobserved components capturing elements of team chemistry. The first component identifies what we call *character players*, or players who positively influence their teammates regardless of the team that they play for; while the second component accounts for the role that a team’s field and front office staff have on team performance to isolate what we call *team players*. This second component also makes it possible to capture a team’s historical ability to consistently turn individual player talents into extraordinary team outcomes, allowing for a relative ranking of MLB teams that can be used to measure organizations on the dimension of team chemistry, or what we refer to as *organizational culture*.

Our methodology also has a natural connection to network statistics that allows us to construct refinements of *fWAR* which isolate a player’s own contribution to team wins irrespective of his teammates, *fWAR*\(^-\), and his contribution adjusted for his effect on his teammates, *fWAR*\(^+\). Using *fWAR*\(^-\), we demonstrate that roughly 44% of the unexplained variation in team wins by *fWAR* is explained by team chemistry. We refer to this total network effect of a team’s players as *tcWAR*, or *team chemistry WAR*, and provide examples of over- and under-achieving teams in recent seasons. Similarly, using *fWAR*\(^+\), we show that *fWAR* tends to overvalue the contribution of low impact players and undervalue the contributions of high impact players to team performance. A player’s net impact on his teammates, i.e. *fWAR*\(^+\) – *fWAR*, is then what we refer to as his *pcWAR*, or *player chemistry WAR*. Our analysis of *pcWAR* confirms the conventional wisdom that star players tend to make their teammates better.

To provide convenient “rules-of-thumb” for general managers in order to maximize team chemistry in roster construction, we next construct age-position profiles for *pcWAR* conditional on player and team characteristics. For example, we show that the conventional wisdom that older players make for good teammates has support empirically, but that the rate of development of team chemistry-related skills varies substantially by position. Our profiles also allow for the estimation of a player’s chemistry *Intangibles*, defined by whether or not their *pcWAR* exceeds or falls short of their profile. Using this measure, we quantify the “David Ross Effect,” so-named after the back-up catcher who we show outperformed his team chemistry profile for much of his career.

The identification of players such as David Ross represents a potential source of competitive advantage for MLB teams. Using player salary data, we show that MLB teams have in the past inconsistently valued the team chemistry-related skills that we capture in our *pcWAR* metric. Only during a player’s free agency years does his compensation positively reflect on average his contribution to team chemistry after controlling
for various other individual factors such as his *fWAR*, age, and experience. Furthermore, MLB teams have placed too low of a value on the *Intangibles* element of *pcWAR* than the value of a win in MLB would suggest is appropriate. One possible explanation for this would be an inability to identify and measure this element of team chemistry, a feature which our analysis overcomes.

The remainder of this paper proceeds as follows: Section 2 provides a brief summary of the relevant literature. Section 3 describes our methodology for measuring team chemistry. Section 4 then details our refinements of *fWAR*, and section 5 presents rules-of-thumb for roster construction. Section 6 concludes and offers some possible extensions of our methodology.

## 2 Literature Review

Accurately measuring team chemistry has in the past been referred to as the “holy grail” of performance analytics (Schrage, 2014). Unsurprisingly, then, a number of other researchers have already made attempts to define and measure chemistry as it relates to team performance in MLB. Their efforts have often focused on identifying the particular traits that denote good “clubhouse culture,” and how this translates into success on the field. Levine (2015) suggests that the presence of a charismatic leader on a roster could have an outsized effect on the performance of his teammates. Similarly, Phillips (2014) uses a regression model to predict that team chemistry can account for up to four wins in a regular season based on characteristics of roster composition like wage parity and demographic variation. Carleton (2013) focuses on two particular players, Brandon Inge and Jonny Gomes, who have been suggested as “good chemistry” players by teammates. He attempts to isolate whether their roster presence affected their teammates’ productivity relative to their expectation for a variety of performance measures. Others have even suggested physiological underpinnings to the chemistry exhibited in the interactions of teammates (SyncStrength, 2016).

In contrast, economists have generally focused more on the particular mechanisms that may generate chemistry spillovers between teammates, framing the problem as one of players serving as complementarity inputs in the production of team wins. The degree of complementarity across players varies substantially across the major professional sports leagues. On one extreme, basketball is a sport where “star” players often have the ability to substitute for their less talented teammates. To this fact, only two of the top ten players as measured by Hollinger’s individual PER metric for the 2016-17 season played for teams that did not make the playoffs. On the other extreme, football presents itself as the quintessential team sport, as it requires more players coordinating their efforts on the field of play.¹ Baseball, on the other hand, seems

---

¹For example, as good as Tom Brady was for the Super Bowl winning Patriots, arguably the defining play in this year’s Super Bowl came while he was not even on the field, and instead when the only quarterback with a higher rating that year was. Matt Ryan, the quarterback for the losing team in the Super Bowl, had a QB rating of 117.1 compared to Tom Brady’s...
to fall somewhere in the middle, with some observers noting its largely individualized nature and others highlighting the importance of offensive and defensive interactions. For instance, Gould and Winter (2009) find that the performance of batters increases with that of other batters on a team. Arcidiacono et al. (2017) suggest that this may be the case because pitchers tend to throw fewer balls to avoid a walk based on the hitting ability of subsequent batters. On the defensive side of the ball, Willis (2017) provides the example of a strong fielding shortstop that may produce greater value to his team if its pitching staff tends to induce ground balls from opposing batters.²

The primary difficulty that others have faced when trying to measure team chemistry in MLB has been their focus on identifying a priori the factors that drive it. Here, we take a novel approach to measuring chemistry by estimating the effect of player complementarities on team wins. Our methodology is designed to look for correlated “mistakes” in the relationship between team wins and the wins-above-replacement metrics of teammates which can tell us something about the complementarities inherent to roster construction. We view this as being consistent with the conventional wisdom that team chemistry is anything that makes teams better than they otherwise would be as purely substitutable individuals. We recognize, however, that this may not be how others view team chemistry. For instance, our analysis is based on the on-the-field interactions of players and not necessarily the off-the-field interactions that tend to get characterized as “clubhouse chemistry.” Insofar as the latter are also reflected in a team winning more games than its collective individual performances would suggest, then they may also be captured by our methodology. Furthermore, if some facet of “clubhouse chemistry” does not lead to either a team outperforming its individual player performances or is manifested only through individual performances already captured by wins-above-replacement, then it has limited to no scope anyway in distinctly explaining the large disparities observed in team performances.

The strength of using FanGraphs particular wins-above-replacement (fWAR) metric for this purpose is its comprehensive nature: It compresses all of the things that a player can do to help his team win both at the plate and in the field or on the mound into one number. We concede, however, that fWAR is not perfect, nor is it alone in this regard.³ There are other well-known alternatives to the FanGraphs methodology (e.g. bWAR produced by Baseball Reference) that can in some extreme cases lead to quite different valuations of a player’s worth. Both FanGraphs and Baseball Reference, however, have taken steps to standardize their calculations such that the definition of a “replacement level” player is the same across both methods (Cameron, 2013). As Miller (2016) notes, the remaining differences in methodology lie in the more subjective year-end QB rating of 112.2. Arguably, one of the most pivotal plays in one of the most historic comeback wins in football history came when Matt Ryan was sacked by the Patriot’s defense and lost the football on a critical third down play in the fourth quarter.²

²Similar frameworks have also been used outside of sports; for instance, in capturing how spatial input-output relationships generate productivity co-movement across sectors of the U.S. economy (Conley and Dupor, 2003).
³FanGraphs (2016b) provides a summary of potential shortcomings.
choices necessary to make the type of comprehensive valuations to which wins-above-replacement aspires, such as whether a pitcher’s quality should be reduced to the outcomes (i.e. runs) for which he is ostensibly responsible or if it should take into consideration how luck and fielding quality may influence these outcomes.

Rather than focus on the details of these differences in methodology, most criticisms of fWAR instead center around the general wins-above-replacement paradigm of translating expected runs into wins in a way that ignores how context may affect team outcomes – for example, a bases-loaded single counts the same as one with two outs and no men on. Such critiques hew closely to common conventions regarding “clutchness,” or whether or not certain players are more capable than others in high leverage situations. Sullivan (2015) notes that while little evidence exists for within-season variation of “clutch” performances being driven by particularly capable teams or players, timing can still be an important factor in explaining how teams may outperform their expectation based on metrics like fWAR. In this spirit, it is worth emphasizing then that our particular use of fWAR is intricately linked to a similar notion that individual players’ performances cannot be measured in isolation, and instead often depend critically on the performance of the players around them. This stands in contrast to the other vein of criticism that wins-above-replacement contains some glaring omission of an activity a player engages in to contribute to a team’s success (e.g. not incorporating stolen bases). While our analysis would surely be influenced by shortcomings like the latter, its ultimate goal is in uncovering shortcomings like the former. We are not interested, however, in the context of how individual play impacts team performance, but instead the correlated nature of performances arising from on-the-field interactions, or team chemistry.

3 Measuring Team Chemistry

In this section, we provide a formal definition for team chemistry in MLB and outline our methodology for measuring it. Our first step along this path is to construct player productivity residuals capturing the difference between the expected number of team wins arising from a player’s performance relative to how many games that player’s team actually won. To measure a player’s performance, we make use of FanGraphs’ wins-above-replacement metric, fWAR, an advanced sabermetric that captures how many total wins a player contributes to his team above a replacement level player at the same position. With these measures in hand, we then move to modeling performance interactions between teammates, or player complementarities, and the effect that they have on team performance.

---

4Details on the data and their sources can be found in the Appendix.
3.1 Player Complementarity as Chemistry

To demonstrate what we mean by team chemistry, consider the archetypal relationships on a baseball team displayed in figure 1. For each panel, we display—through several contour lines—the mix of player talents across two particular positions on a baseball team and their hypothesized impact on team wins (where darker lines moving towards the northeast signify increases in team performance). On any contour line, the slope at a particular point denotes the increase (decrease) in fWAR required at one position to displace a loss (increase) in fWAR at the other position in order to hold overall team performance constant. The degree to which players are substitutable or complementary determines the curvature of these contour lines.

We propose that the stronger complementarities are among players the greater the scope that exists for team performances to be differentiated on the dimension of chemistry. In the top left panel, we display what this relationship might look like for a designated hitter (DH) and a starting pitcher (SP) on the same team. Given that these two types of players will never find themselves on the field at the same time, and also engage in extreme types of activities, it’s natural to imagine that their performances are perfectly substitutable, consistent with the linear contour lines in this panel. In other words, a unit increase in wins-above-replacement from a team’s DH should be able to perfectly substitute for a unit decrease in wins-above-replacement from a starting pitcher. This sort of neutral interaction among players mirrors the implicit assumption underlying the construction of fWAR values for teammates and how they map in the aggregate into a team’s fWAR.

Now, consider the relationship between the 3rd and 4th place hitters in a team’s lineup, as displayed in the top right panel of the figure. For the 3rd and 4th hitters in a lineup, it is much more natural to assume that their performances are intricately linked when determining team performance. As good as the 3rd hitter might be (e.g. batting .300 and consistently getting on base), without a 4th hitter to either drive him in to score runs or protect the 3rd hitter from getting intentionally walked in pivotal hitting situations (e.g. runners on base with two outs), the team’s performance is likely to suffer. The complementary nature of these two players’ performances in dictating overall team performance is captured by the strong kinks in the contours displayed in this panel. These sorts of complementarities are very certainly lost in the construction of fWAR at the individual level and, thus, also by aggregating to the team level.

Of course, the 3rd and 4th hitter might be an extreme case as well. In the bottom panel, we instead display what might be the appropriate degree of complementarity between the second basemen (2B) and shortstop (SS). Here, one could imagine that some amount of substitution exists between these defensive positions: a shortstop with incredible range might be able to cover up for a slow-footed second baseman in fielding ground balls up the middle. Alternatively, it would be natural to imagine that the two positions also have a degree of
complementarity in that both of these players are also necessary for a team to successfully turn a double play. Furthermore, depending on where these two positions bat in the lineup, further offensive interdependencies may also come into play. Importantly, however, even the more moderate degree of interdependency displayed in this panel would not be captured in the construction of $fWAR$.

### 3.2 $fWAR$ and Team Wins

Instead of building a structural model of how individual players contribute to the different outcomes through the course of a game and then how these outcomes translate into the likelihood of a team winning a game, we take a more statistical approach. Specifically, we search for systematic correlations in the misspecifications of $fWAR$ across teammates by incorporating into our analysis how a team actually performed. If these misspecifications are systematic both across teams and individual players as teammate relationships change, our statistical model will attribute them to the player complementarities that must exist between teammates along the lines of what is described in figure 1.

We show here that this tends to manifest itself in the fact that simply summing the $fWAR$ values for a team across its players does not perfectly replicate its wins above those expected of a team comprised entirely of replacement-level players. To get a sense of exactly how important player interactions may be to team performance, we regressed the number of wins for each team on the sum total of its players’ $fWAR$. Specifically, we ran a linear regression of the form

$$W_{nt} = \alpha + \beta fWAR_{nt} + \varepsilon_{nt},$$

where $W_{nt}$ is the number of wins of team $n$ in season $t$ and $fWAR_{nt}$ is the sum total of FanGraphs’ wins-above-replacement statistics for all players on team $n$ in season $t$.

The $\varepsilon_{nt}$ in this regression are what we call team productivity residuals. We refer to them as such because in many ways they represent the baseball equivalent of the famous “Solow residual” used in economics to measure the productivity of firms. An MLB team with a positive $\varepsilon_{nt}$ was a team who out-performed, or won more games than what could be attributed to the sum of its individual player performances (or in economic terms, a firm that produced more output than the usage of its individual inputs would suggest). Alternatively, a team with a negative residual would be a team who despite perhaps having a number of strong individual performances (as measured by $fWAR$) under-performed as it pertains to wins.

The results from this regression using MLB team data from the 1998-2016 seasons, shown in table 1, conducted a similar analysis in his examination of $fWAR$. See Solow (1957) for details on the Solow residual.
provide several insights. First, it is clear that the estimate of $\beta$ is very close to 1.\footnote{In fact, the null hypothesis of $\beta = 1$ cannot be rejected at any standard confidence levels.} This is intuitive given how $fWAR$ is constructed \cite{FanGraphs}, but also allows us to confidently use the idea that increasing a team’s $fWAR$ should have a one-to-one relationship with their number of wins. Furthermore, the estimate for $\alpha$ rounds to 48. This estimate, too, has a natural interpretation of being the number of wins one would expect a team full of replacement level players to accrue. At 48, clearly a team with only replacement level players is far from an average, or 0.500 winning percentage, team. With that being said, it is consistent with the construction of $fWAR$; and, thus, serves as a benchmark for us to evaluate teams.

The team productivity residuals, $\hat{\varepsilon}_{nt}$, are our estimate of the element of team performance that is unexplained by the sum of its players’ individual performances, and the variation that we may potentially attribute to a team’s chemistry. Based on the $R^2$ value of the regression, this amounts to about 20% of the variation in team wins in our sample. The left-hand side panel of figure 2 further demonstrates just how important this element is by plotting a kernel density function of $\hat{\varepsilon}_{nt}$ (blue line). With a standard deviation of 5.2 wins and a range equal to approximately 40 wins, it is evident that a considerable portion of the variability in team performance cannot be explained by $fWAR$. Our contention is that at least part of this variation is the result of attempting to construct an individualized measure of performance while abstracting from the complex set of interdependencies amongst teammates.

To see why, consider the following interdependencies captured in our regression. In constructing a roster, teams face a problem of maximizing wins, $W$, subject to a payroll constraint and MLB regulations like the luxury tax. Suppose this production function, $w$, takes the form

$$W = w[d(f,p), l(b)],$$

where $w$ takes as its inputs a team’s defensive and offensive production. Defensive production is captured by $d(f,p)$, a function that describes how fielding ($f$) and pitching ($p$) resources determine defensive value; and offensive production is captured by $l(b)$, which specifies how batters ($b$) determine offensive value given a lineup configuration. If the team’s defensive ability depends upon the specific mix of fielding and pitching inputs, or if teams maximize the returns to their batter’s output based on their order in the lineup, then team chemistry will be evident in the degree of substitution/complementarity between inputs.

We do not observe the functional form $w$ takes for each team, but we do have the extensive work of sabermetricians to appeal to on this matter. In fact, the construction of $fWAR$ resembles this production function in many ways, as it incorporates fielding, pitching, and hitting metrics separately by converting them to run-equivalent values for each player which are then translated to team win values by using an historical
run differential-win relationship. If the technology for turning player talents into team wins is linear with respect to the sum of its players’ individual $fWAR$, i.e. if player performances are perfect substitutes, then the relationship below should hold and the residuals of our regression should be zero.

$$W = \sum_i fWAR_{int} + \alpha. \quad (3)$$

However, if interactions exist between teammates causing this relationship to be nonlinear, then our regression will be misspecified. When we replace $fWAR$ in our regression with our measure that adjusts for the strength of teammate interactions, $fWAR^{-}$, this is exactly what we see. In this case, our estimate of $\beta$ is just a little higher than 1 while our estimate of $\alpha$ still rounds to 48 wins, as seen in table 1. In addition, based on the higher $R^2$ value of this regression, accounting for player performance interactions reduces the unexplained variation in team wins by close to 44%.

The result of this adjustment on our team productivity residuals can be seen in the left-hand side panel of figure 2. The kernel density of team productivity residuals using $fWAR^{-}$ (red line) is much more concentrated than before with a standard deviation of 3.9 wins and a range of about 30 wins. The right-hand side panel of figure 2 then provides confirmatory evidence that this result does in fact likely stem from player complementarities by scattering our unadjusted team productivity residuals against a measure of the concentration of a team’s players’ $fWAR$ in each season, where concentration is captured by the Gini coefficient. The more unequally distributed $fWAR$ is across a team’s players, the higher the Gini coefficient. Consistent with the lower variance of our kernel density estimates based on $fWAR^{-}$, the local cubic polynomial regression fit in the figure demonstrates that a negative relationship exists between a team’s productivity residual and the concentration of its players’ $fWAR$. Taken together, these results suggest that complementarities between individual players, or what we call team chemistry, do indeed have scope for explaining some of the unexplained variation in team performance by $fWAR$.

---

$^8$The Gini coefficient captures the statistical dispersion of a given distribution. It takes a value between 0 and 1, with values closer to 1 denoting more unequal distributions (Ceriani and Verme [2012]). In our calculation, we drop negative $fWAR$ values in order to make our Gini coefficient estimates consistent with the standard interpretation.

$^9$A linear regression fit also yields a statistically significant negative relationship.
3.3 Team Wins and Teammate Interactions

Next, we focus on our methodology for decomposing team productivity residuals into player-specific productivity residuals. The basis for this decomposition is the following identity,

\[ \hat{\varepsilon}_{nt} = \sum_{i} \hat{\varepsilon}_{int} \]

\[ = \sum_{i} \hat{W}_{int} - \sum_{i} fWAR_{int}, \]  

where \( \hat{W}_{int} \) is a measure of the contribution of player \( i \) to team wins such that \( \sum_{i} \hat{W}_{int} = W_{nt} - \alpha \). In order to construct player productivity residuals, we must carefully define a player’s expected contribution to his team’s wins based on his share of team outcomes. First, we construct two weights in order to capture how a player’s placement in the batting lineup \( (l_i) \) and his defensive playing time \( (d_i) \) affect his expected contribution,

\[ l_i = \sum_{j=1}^{9} b_j S_{ij} \]  

\[ d_i = \sum_{j=1}^{8} p_j g_{ij}, \]

where \( S_{ij} \) denotes the number of times player \( i \) appeared in the \( j \)th slot of the lineup and \( g_{ij} \) denotes the number of times player \( i \) appeared at each of the eight non-pitching defensive positions as a share of his total appearances. The adjustment weights, \( b_j \) and \( p_j \), then describe the relative importance weight given to their respective variables. Lineup weights are defined following [Tango et al. (2007)](http://example.com), while player position weights are defined following FanGraphs’ \( fWAR \) methodology [FanGraphs (2016a)](http://example.com).\(^{10}\) We then normalize each weighting scheme to sum to 1, with the resulting values reported in table 2.

With these weights in hand, we then proceed to the construction of \( \hat{W}_{int} \) for each player based on his

\(^{10}\)Lineup weights reflect the likelihood of getting additional plate appearances within a game, rather than any notion of proximity toward other players in the lineup.
position and his share of his team’s players’ appearances ($g_{ip}$):

$$
\hat{W}_{int} = \eta_i g_i W_{nt} \quad (8)
$$

$$
\eta_i = \begin{cases} 
0.57 & \text{if } i \text{ is a position player} \\
0.43 & \text{if } i \text{ is a pitcher} 
\end{cases} \quad (9)
$$

$$
g_i = \begin{cases} 
\frac{l_i * PA_i + d_i * DOuts_i}{\sum_{k \neq i} (l_k * PA_k + d_k * DOuts_k)} & \text{if } i \text{ is a position player} \\
\frac{POuts_i}{\sum_{k \neq i} POuts_k} & \text{if } i \text{ is a pitcher} 
\end{cases} \quad (10)
$$

FanGraphs constructs $fWAR$ such that players contribute 1,000 WAR per 2,430 games league-wide (162 games for 30 teams). The terms 0.57 and 0.43 in the above equation correspond to the proportion of league-wide WAR they apportion to position players and pitchers, respectively. This split is based on the assumption that because position players appear on both sides of the ball, their contribution should be weighted somewhat higher [FanGraphs 2016a]. To generate appearance weights for position players, we use the sum of plate appearances ($PA$) and defensive outs ($DOuts$) differentially weighted by the lineup and position weights derived above and reported in table 2. For pitchers, outs recorded ($POuts$) proves to be the most precise measure for capturing a variety of pitching contributions (middle relievers, one-out guys, etc.). Finally, we scale our appearance measures to allow us to make per-game comparisons across the different measures, such that plate appearances are multiplied by a three appearance per-game scalar and defensive and pitching outs are multiplied by a 27 outs per-game scalar.

When aggregated across players on a given team in a given season, our player productivity residuals measure the difference between a team’s actual win count and what it would be expected to be based on the sum total of individual player performances as measured by $fWAR$. We then model the interactions between player productivity residuals as a spatial autoregression (SAR),

$$
\hat{\varepsilon}_{int} = \rho A \hat{\varepsilon}_{int} + \nu_{int}, \quad (11)
$$

where $A$ is an adjacency matrix identifying teammates in a given season. Typically, an adjacency matrix is a symmetric matrix with 0’s on the diagonal and 1’s off the diagonal “connecting” teammates. However, in order to capture potential interdependencies in teammate relationships which correspond to their positions in the field and lineup, we replace the 1’s with their combined weighted appearances used in the construction of $\hat{W}_{int}$ summed over the number of MLB seasons during which players appeared together on the same team. This allows for both added weight to be given to repeated “connections” in explaining player
performance interactions as well as to be given to more valuable and complementary defensive and lineup position combinations in explaining team performance.\footnote{A detailed discussion of the construction of our adjacency matrix appears in the Appendix.}

Furthermore, we assume that a factor structure exists for the SAR residuals, $\nu_{int}$, such that player productivity residuals are summarized by a player-season specific ($f_{it}$) component, as well as a team specific component ($\lambda_n$). The $f_{it}$ trace out a player’s career arc, potentially across several teams, and reflects whether that player finds himself among over- or under-performing teammates in each season. Identification of this latent variable is, therefore, predicated on roster turnover. Because of this, it will be more difficult in general for us to establish such a player vs./ team breakdown the less roster turnover exists on a team over time. The team specific component, on the other hand, reflects an organization’s average historical tendency to over- or under-perform relative to the collection of its players.

Solving for $\hat{\varepsilon}_{int}$ yields our spatial factor model with spatial weight matrix $W = (I - \rho A)^{-1}$.

$$\hat{\varepsilon}_{int} = (I - \rho A)^{-1}(f_{it}\lambda_n) = W F \Lambda$$

(12)

To estimate this model, we use a two-step estimation procedure described in the Appendix. In the first step, an estimate of $\rho$ is obtained by maximum likelihood conditional on a scale normalization on $A$. Given $\rho$, the factor model is then estimated by spatial principal component analysis (SPCA) to extract the latent player-season and team specific components up to a scale normalization on $\lambda$ (Demsar et al. 2012). In the next section, we provide further motivation for what we aim to capture in these factors in terms of team chemistry.

4 The Network Effects of Team Chemistry

To measure the interdependence of teammates’ performances, we borrow heavily from the social and economic network analysis literature (Jackson 2008). Our spatial factor model fits the definition of a network. The players on a team in a given season make up the “nodes” of the network, with the strength of the connections between teammates summarized by our adjacency matrix, factors, and their loadings. In other words, our model is simply a statistical framework for measuring the importance of correlations across team and teammate performances. In this section, we refine $fWAR$ in order to take into account these correlations; and, at the same time, construct new metrics that can be used to evaluate players’ contributions to team chemistry.
4.1 Sources of Team Chemistry

Our methodology for measuring team chemistry boils down to nothing more than a decomposition of the spatial correlation matrix of teammates’ productivity residuals into an exact linear combination of latent factors. To see this, consider that we can decompose our player productivity residuals into two parts: 1) a part that is unique to each player that we attribute to measurement error in team productivity residuals, and 2) a part that can be explained by each player’s interactions, or spill-overs, with his teammates that we attribute to team chemistry.

\[
\hat{\varepsilon}_{\text{int}} = w_{ii}(f_{it}\lambda_n) + \sum_{j \neq i} w_{ij} (f_{jt}\lambda_n) . \tag{13}
\]

We associate positive spill-overs with “good team chemistry” and negative spill-overs with “bad team chemistry.” We do not take a stance on what drives these spill-overs between teammates; and, in all likelihood, our latent factors probably capture a combination of many of the determinants of team chemistry that others have already explored. However, by not restricting them ex-ante, they likely also embody elements of team chemistry that have not previously been able to be measured.

The extent to which we provide context for our factors is only to appeal to the work of other social scientists who have singled out certain psychological traits, such as “character” and being a “team player,” as being attributes of individuals in groups that excel in working together. By allowing for two factors and restricting their loadings such that \( F = [ch, tp] \) and \( \Lambda = [l, \lambda] \), where \( l \) is a unit vector across teams, we can restrict our factor model to embody similar features.

\[
\hat{\varepsilon}_{\text{int}} = w_{ii} (ch_{it}l_n + tp_{it}\lambda_n) + \sum_{j \neq i} w_{ij} (ch_{jt}l_n + tp_{jt}\lambda_n) \tag{14}
\]

We think of the factor \( ch \) as capturing a player’s innate character, as through this factor players demonstrate spill-overs to their teammates which do not depend on the identity of their team. In contrast, we think of the factor \( tp \) as capturing a player’s contribution that is more closely linked to the “match quality” of his current team (via \( \lambda \)), which we term as the team player factor.

Teams with large \(|\lambda|\) are then said to exhibit good organizational culture, as they either reinforce positive spill-overs (\( tp < 0 \& \lambda > 0 \)) or minimize negative spill-overs (\( tp > 0 \& \lambda < 0 \)) between teammates. Figure 3 plots rankings from zero to 100 for all 30 MLB teams across the 1998-2016 seasons based on our estimated values of \(|\lambda|\). Certain organizations stand out along this dimension. For instance, the St. Louis Cardinals and San Francisco Giants are the top two teams in our ranking; while others do not fair nearly as well, such
as the Chicago Cubs and White Sox. It is important to keep in mind, however, that these rankings represent an average for the 1998-2016 seasons and are potentially subject to change over time.

4.2 Adjusting fWAR for Team Chemistry

If fWAR measurements are indeed influenced by teammate interactions, then the regression underlying our team productivity residuals is misspecified. Namely, fWAR may be under- or over-counting the importance of individual contributions to team wins by ignoring the interactions between teammates. To adjust for this possible source of bias, we construct an alternative measure called fWAR− which subtracts from the fWAR of each player the portion of his productivity residual that can be explained by his teammates' residuals. In network statistics, this is often referred to as the “in-degree” for a node.

\[
fWAR_{int}^- = fWAR_{int} - \sum_{j \neq i} w_{ij} f_{jt} \lambda_n
\]  \hspace{2cm} (15)

Recall that figure 2 demonstrated the relative importance of adjusting fWAR in this way for explaining deviations of team productivity residuals from zero. We can get a sense of the impact that this adjustment has on the productivity residual for any individual team by examining the aggregation of the differences between fWAR and fWAR− over teammates in each season. This is often referred to as the network’s “total-degree.” We call it “team chemistry wins-above-replacement,” or tcWAR, and scale it by −β from the above regression so that we can relate it directly to figure 2.

\[
tcWAR_{int} = -\beta \sum_i \sum_{j \neq i} w_{ij} f_{jt} \lambda_n
\]  \hspace{2cm} (16)

Figure 4 scatters a team’s productivity residual in each season against its tcWAR. The figure is constructed so that the x-axis coordinate (tcWAR) is equal to the number of team wins (y-axis coordinate) explained by team chemistry. Some of the best and worst teams on both ends of the chemistry spectrum are noted in the figure. For instance, of the 2012 Orioles’ nearly 15 team wins above fWAR’s expectation, tcWAR attributes roughly 5 of these to good team chemistry. In contrast, of the nearly 15 wins below expectation of the 1998 Mariners, tcWAR attributes roughly 8 to bad team chemistry.

Similarly, we can refine fWAR as a measure of player performance by taking into account how much a player affects his teammates’ performances. Here, we add to fWAR− the contribution of each player to all of his teammates’ productivity residuals, or what is referred to in network statistics as the “out-degree” of
a node. We call this measure $fWAR^+$. 

$$fWAR^+_{int} = fWAR_{int} + \sum_{i \neq j} w_{ji} f_{it} \lambda_n \leftarrow \text{“Out-degree”}$$  \hspace{1cm} (17)

Figure 5 scatters $fWAR^-$ and $fWAR^+$ versus $fWAR$. Interestingly, $fWAR^-$ and $fWAR$ on an individual player-season basis are very highly correlated, with the plotted points clustered fairly closely around the 45 degree line. Thus, it is the aggregation of somewhat small differences at the player level that leads to the drastic reduction in the unexplained variance of team performance in figure 2. For $fWAR^+$, on the other hand, the differences are much more pronounced. In particular, our analysis suggests that $fWAR$ overestimates the relative performance of low impact ($fWAR \leq 1$) and underestimates the relative performance of high impact ($fWAR \geq 4$) players on team performance.\footnote{One way in which this result could manifest itself is if low/high $fWAR$ players tend to accrue a disproportionate share of their $fWAR$ during non-pivotal/pivotal game situations.}

4.3 Player Evaluation

The difference between $fWAR^+$ and $fWAR$ can be used to evaluate players on the basis of their contribution to team performance through their impact on their teammates. In network statistics, this is what is called the “net-degree” for each node.

$$pcWAR_{int} = \sum_{i \neq j} w_{ji} f_{it} \lambda_n - \sum_{j \neq i} w_{ij} f_{jt} \lambda_n \leftarrow \text{“Net-degree”}$$ \hspace{1cm} (18)

In keeping with our terminology above, we instead refer to it as “player chemistry wins-above-replacement,” or $pcWAR$. In figure 6, we plot the $pcWAR$ for all player-season combinations in our dataset relative to a player’s $fWAR$. Notice that summing a player’s $pcWAR$ and $fWAR$ reproduces our $fWAR^+$ metric, such that adding the x-axis and y-axis coordinates for each player-season in this figure provides a sense of his true value to his team.

The conventional wisdom that good players make their teammates better is confirmed by our analysis of $pcWAR$, as figure 6 demonstrates a strong positive correlation exists between $pcWAR$ and $fWAR$ for all player-season combinations in our sample. The vertical lines in the figure correspond to thresholds for $fWAR$ used by FanGraphs to distinguish Good from Star players ($fWAR = 4$) and Scrub from Role players ($fWAR = 1$). Star players tend to add anywhere from about 0 to 1.5 wins to their team through their indirect impact on the performance of their teammates, whereas Scrub players tend to add from -0.5
to 0 wins to their team. In between, there exists considerable variation with players contributing from -0.5 to 0.5 wins through team chemistry.

Figure 7 ranks all active players through the 2016 season on the basis of their career average \(pcW AR\) values. The left-hand panel of the figure shows the top 25% of players on this dimension, while the right-hand panel shows the bottom 25%. Many of the top players in the game dominate our leaderboard, with Mike Trout the undisputed champion in this regard, averaging over one-half win of additional value through team chemistry over the course of his career. The rest of the top four is comprised of Clayton Kershaw, Kris Bryant, and Joey Votto, who each have contributed about 0.4 wins on average for their careers via team chemistry. At the bottom of the rankings, only three players on average have subtracted above 0.2 wins from chemistry: Joey Rickard, Tyler Goeddel, and Yasmany Tomas.

While our estimates for \(pcW AR\) may seem small at first glance in terms of win value, they are of a non-trivial economic value. With a team win valued at roughly $6 million in MLB, the value of team chemistry alone for some of the game’s best players is just as high according to our \(pcW AR\) metric as what \(fW AR\) would assign to a typical role player on the team [Cameron 2014]. In fact, even a player whose \(fW AR\) was 0 and \(pcW AR\) was as low as 0.1 would still be worth paying the MLB minimum salary.

5 Chemistry and Roster Construction

Our aim in this section is to develop some convenient “rules-of-thumb” for MLB general managers to follow when considering team chemistry in roster construction. We first explore the drivers of player chemistry by constructing age-position profiles for \(pcW AR\) conditional on player and team characteristics. These profiles then allow us to rank players on the dimension of their unobserved chemistry \(Intangibles\) and quantify the “David Ross Effect.” Because salary negotiation plays such an important role in roster construction, this leads naturally then to a discussion of the value of team chemistry.

5.1 Age-Position Profiles and Intangibles

We construct conditional average age-position profiles for team chemistry from the following regression,

\[
pcW AR_{it} = \sum_p \gamma_p(pos_{p_{it}}) + \sum_p \theta_p(pos_{p_{it}} \cdot age_{p_{it}}) + \sum_p \psi_p(pos_{p_{it}} \cdot age_{p_{it}}^2) + \sum_p \tau_p(pos_{p_{it}} \cdot age_{p_{it}}^3) + \sum_p \omega_p(pos_{p_{it}} \cdot age_{p_{it}}^4) + \sum_k \delta_k X_{k_{it}} + \sum_h \phi_h Z_{hit} + \xi_{it},
\] (19)
where $pos$ is an indicator variable for a players’ primary defensive position, including the designated hitter and a “utility” category for players who tend to play multiple defensive positions, $age$ is a player’s age, $X$ is a vector of player characteristics including $fWAR$ and controls for MLB and team games played and $Z$ is a vector of league, team, and manager indicator variables.

Figure 8 plots our conditional average age-position $pcWAR$ profiles with 95% confidence intervals. The conventional wisdom that older players make for better teammates is certainly consistent with these profiles, as they tend to slope upward with age on average across all positions. Some additional interesting patterns also emerge from this analysis. For instance, designated hitters and corner infielders and outfielders have the steepest profiles, with $pcWAR$ values that are on average positive starting in a player’s early twenties. Of the remaining positions, catchers, relief pitchers, and utility players demonstrate the steepest profiles, with conditional $pcWAR$ values that are on average positive starting in a player’s early thirties. Only the profile of starting pitchers is statistically insignificant across all ages.\(^{13}\)

By conditioning our age-position profiles on so many observable dimensions, we can also isolate the player Intangibles of team chemistry. In other words, we can measure the individual contributions to team wins through chemistry that are not associated with a player’s $fWAR$, his team or manager, as well as his age or experience. The estimated coefficients of the above regression demonstrate that many of these factors are indeed important predictors of team chemistry. For instance, one additional win-above-replacement increases a player’s $pcWAR$ by a statistically significant 0.09 wins on average (see the table in the Appendix). We use the residuals, $\xi_{it}$, from our conditional age-position profile regressions to rank active players through the 2016 season on their career average Intangibles. Positive residuals capture players whose contributions to team chemistry exceed their conditional age-position profile, whereas negative residuals correspond to players who fall short of their profile. Figure 9 displays our rankings, where the left-hand panel shows the top 25% of players on this dimension and the right-hand panel shows the bottom 25%.

Our Intangibles rankings are interesting for a number of reasons. First, our top four players are now very different than they were for $pcWAR$. Kevin Keirmaier is the active leader on this dimension of team chemistry with an average contribution of a little more than 0.1 wins coming from his Intangibles. Just below him are Giancarlo Stanton, Joey Votto, and the recently deceased Jose Fernandez. Recall that of these four, only Joey Votto was also in the top four of our $pcWAR$ rankings. Furthermore, one of Votto’s peers in our top four $pcWAR$ rankings (Kris Bryant) ends up in the bottom ten in our Intangibles rankings; and he is not alone in terms of star power, with players such as Matt Carpenter and Dexter Fowler also appearing near the bottom of our rankings where Nori Aoki and Nomar Mazzara share the bottom spot.

\(^{13}\)We want to caution anyone from taking the results from this regression as “causal” estimates of age on intangibles, as the estimated coefficient is most likely also confounding a selection effect for older players. In other words, having good intangibles may make it more likely for a player to remain in the game for longer.
5.2 The “David Ross Effect”

At this point, a word of caution is warranted. Many of the players that we find have negative \textit{Intangibles} are the same type of player that any MLB franchise would be happy to build their team around. In other words, depending on a player’s position, his age, and current manager, etc. he could still have a strong positive influence on his teammates even despite a negative \textit{Intangibles} measure. A more appropriate interpretation of our rankings is then that they provide an indication of those players who have a knack for exceeding expectations on the dimension of team chemistry. We call this the “David Ross Effect.”

The esteem with which the 2016 World Champion Chicago Cubs held their teammate David Ross and his contribution to their success has by now become well known. The ability of a team to identify players like him is, therefore, a potential source of competitive advantage that is made possible by our team chemistry metrics. What makes David Ross uniquely suited to our analysis is that as a back-up catcher his $fWAR$ defines him as a role player; but, as a teammate, he is routinely characterized as someone who makes everyone around him better. We are able to provide evidence to support these claims with our metrics. Figure 10 plots the $pcWAR$ and \textit{Intangibles} values for David Ross through the 2016 season against his $fWAR$ values. More than one labeled instance of a season occurs whenever he was traded. For most of his career, and across several different teams, David Ross exhibited the sort of beneficial relationship with his teammates that his reputation attests to, evidenced by his mostly positive $pcWAR$ values. Furthermore, his \textit{Intangibles} reveal a player who consistently outperformed his age-position profile even at low levels of $fWAR$ and with limited playing time.

Players such as David Ross are true “diamonds-in-the-rough” according to our analysis, with their full impact on team performance likely to fly under the radar according to traditional performance metrics. This raises the question of who is likely to assume his mantle for the Cubs. Interestingly, our estimates suggest an embarrassment of riches in this regard, with a number of players on the Cubs 2017 roster that have positive career average \textit{Intangibles}. These include star players such as Jake Arrieta, Anthony Rizzo, Kyle Schwarber, and Ben Zobrist, but also a number of role players like Albert Almora, Hector Rondon, Pedro Strop, Tommy La Stella, and Justin Grimm.

Given how rare that we find that this type of player is in our analysis, one might expect then that MLB teams would be willing to pay a premium for their services. For example, others have already documented the importance of wins-above-replacement in pricing player services in MLB\textsuperscript{14}. For this reason, we might expect our $pcWAR$ metric to also be priced into player compensation. Whether or not this extends to a player’s \textit{Intangibles}, which are arguably much more difficult to observe than age, position, and the other

\textsuperscript{14}See Cameron \textsuperscript{2014} and Paine \textsuperscript{2015}.
variables that we condition, is unclear.

### 5.3 Putting a Price on Team Chemistry

To see how MLB teams have historically valued chemistry-related skills, we run a player-level regression of log annual salary adjusted for inflation as our dependent variable on a player’s career $fW AR$ and $pcW AR$.\(^\text{15}\)

This regression takes the following form,

$$
\log(\text{salary})_{it} = \sum_{c=0}^{2} \gamma_c \sum_{n=1}^{t-1} fW AR_{in} + \sum_{c=0}^{2} \beta_c \sum_{n=1}^{t-1} pcW AR_{in} + \sum_{c=0}^{2} \theta_c FA_c + \sum_{p} \lambda_p (pos_{it} \times mlbExp_{it}) + \sum_{p} \tau_p (pos_{it} \times mlbExp_{it}^2) + \sum_{p} \phi_p (pos_{it} \times age_{it}) + \sum_{n=1}^{n} \psi_n teamExp_{it} + \alpha_i + \epsilon_{it}, \tag{20}
$$

where $pos$ is an indicator variable for a players’ primary defensive position, including the designated hitter and a “utility” category for players who tend to play multiple defensive positions; $age$ is a player’s age on January 1\(^\text{st}\) of the year in which season $t$ occurs; $mlbExp$ is the number of MLB games in which the player has appeared; $teamExp$ indicates the number of games the player has appeared in with his current team; $FA$ is an indicator variable which takes on one of three values denoting whether a player is in the pre-arbitration (0-2 years of service), arbitration-eligible (3-5 years), or free agent-eligible (6+ years) stage of his career; and $\alpha$ is a player fixed effect.

The use of a player fixed effect focuses our analysis on the variation “within” player salary histories. For this reason, we restrict our sample of players to only those with careers that began at some point during the 1998-2016 seasons. The interaction terms with age and experience then serve to capture any nonlinear variation in how players’ career earnings trajectories vary across defensive positions. As we will soon explain, the inclusion of the $FA$ indicator variable serves to capture important differences in how player salaries are determined throughout a career based on the labor market structure of MLB and changes in the bargaining power of players according to service time.\(^\text{16}\)

We are primarily interested in obtaining estimates for $\gamma$ and $\beta$, the coefficients on career $fW AR$ and $pcW AR$, respectively. Interacting these variables with our service time-status indicator allows us to estimate how these skills may be differentially priced over a player’s career. As the first column of table 3 shows,

\(^{15}\)The functional form of these regressions is similar in spirit to the wage determination model presented in Mincer (2008).

\(^{16}\)A year of service is defined as 172 days during a season on an MLB roster. Because we don’t observe this number directly, we approximate these thresholds by simply counting the number of seasons in which each player appears in our sample. This means it is possible that some arbitration-eligible players are counted as pre-arbitration or as free agent-eligible. However, our regression results are qualitatively similar when we use game appearances as a proxy for whether a player completed a year of service in this calculation.
cumulative \( fWAR \) is indeed priced differentially throughout a player’s career. During the pre-arbitration stage of a player’s career, a one win-above-replacement increase in career \( fWAR \) through the previous season leads, on average, to a statistically significant earnings increase in the current season of approximately 9 percent. For a player in the arbitration-eligible portion of his career, this increases slightly to 12 percent. Finally, players with six or more years of MLB service time see an average increase in earnings of only 4 percent for each additional unit of career wins-above-replacement.

To understand the pricing pattern demonstrated in this result, it is useful to consider the bargaining position of the player. The pre-arbitration period corresponds to a player’s first three years of service time, measured by days spent on the 25-man roster of any MLB team. Unless released or traded, players are bound to the team that drafted them during this period. The vast majority of these players earn either a minimum salary determined by collective bargaining between MLB and the MLB Players Association or a somewhat higher salary on a season-by-season basis that is at the discretion of the team. Performance and salary are therefore likely to be only somewhat correlated during this time. Furthermore, even if a player were to sign a long-term contract during this time, they lack the bargaining power they would have if their services were being priced by the entire league, an economic situation referred to as monopsony.

If a player has still not signed a long-term contract after three years of service, they become eligible for salary arbitration, whereby the player and team submit proposed salaries to an independent third party that makes a binding determination on the player’s salary, largely on the basis of similar player performances.\footnote{While three years is the general cut-off for salary arbitration, players that are in the top 22% of service time among those with more than two but less than three years of service become eligible for arbitration a year early. This “Super Two” cutoff is designed to prevent teams from delaying a player’s call up from the minor leagues by a few weeks to avoid salary arbitration. [FanGraphs 2017]} Though players still have limited bargaining power during this period, the slight increase in the return to cumulative \( fWAR \) that we observe is consistent with their improved bargaining position afforded by the arbitration process. When a player has not signed a long-term contract after accruing six or more years of MLB service time, he becomes eligible to sign with any team as a free agent.

Once a player enters free agency, it is much more common for him to sign a multi-year contract. Multi-year contracts add a further complication to our regression, since their pricing reflects a combination of both past and expected future performances. This could largely explain the smaller coefficient that we find on cumulative \( fWAR \) during free agency. However, the competitive landscape of free agency may also force teams to consider a broader range of factors as they submit contract offers to players. In fact, our estimated regression coefficients on cumulative \( pcWAR \) suggest that a player’s chemistry-related skills are perhaps one of the additional things considered.

The first column of table 3 shows that cumulative \( pcWAR \) is also priced differentially throughout a
player’s career. The effect on earnings of a one win-above replacement increase in career pcWAR is large, negative and statistically significant in the pre-arbitration and arbitration-eligible periods. This suggests that the lack of competitive pressure in these years allows teams to avoid compensating players for their chemistry-related skills. However, this effect reverses once a player becomes eligible for free agency, as the same increase in pcWAR now leads to a statistically significant salary gain of roughly 9.5 percent.

It may seem counterintuitive that the marginal return to a unit increase in cumulative pcWAR is higher than that for a unit increase in cumulative fWAR during free agency. However, as a relatively scarce resource (its standard deviation is nearly 15 times smaller than cumulative fWAR), it makes sense that cumulative pcWAR is priced higher on the margin. That said, it could also just as easily be the case that what we find reflects a team’s pricing of some of pcWAR’s underlying correlates. For instance, if MLB teams follow the conventional wisdom that high-performing players will have the biggest chemistry effects, they may simply pay more for individuals who they expect will rank highly in the future on metrics such as fWAR. The larger coefficient on cumulative pcWAR would then reflect this fact.

To investigate this possibility, we run an alternative earnings regression that instead considers separately our Intangibles measure (i.e. $\xi$) and the observed component of pcWAR (i.e. pcWAR $-\xi$). These results, presented in the second column of table 3, show that the two individual components of cumulative pcWAR are indeed priced differently in free agency (i.e. their coefficients have opposite signs). A player with positive career Intangibles would be undervalued relative to his contribution to the team; and, conversely a player with negative career Intangibles would be overvalued. In contrast, the observed component of cumulative pcWAR for either player would still be appropriately valued. This result suggests an important application for our metrics. Though teams appear to value their players’ chemistry-related skills, they clearly misprice their Intangibles, most likely because they are not easily observed.

6 Conclusion

In this paper, we outlined a methodology for quantifying how a player may influence his team’s performance outside of his direct contribution measured by advanced individual performance metrics like wins-above-replacement. We introduced in the process fWAR$^-$, fWAR$^+$, tcWAR, and pcWAR as new advanced metrics that quantify the indirect effects of players on their teammates and team performance while providing an intuitive analog to FanGraph’s well-documented fWAR metric. With these new metrics, we then outlined the importance of accounting for player interactions in explaining team performance differentials unexplained by fWAR.

We also developed convenient “rules-of-thumb” for general managers to follow when considering team
chemistry in roster construction, demonstrating that star and older players make for good teammates and that the rate of development of chemistry-related skills varies by position. We were able to isolate a player’s chemistry *Intangibles*, defined by whether or not their *peWAR* value exceeds or falls short of their age-position profile conditional on player and team characteristics. This allowed us to quantify what we call the “David Ross Effect,” so named after the back-up catcher who consistently outperformed his profile for much of his career. Furthermore, we demonstrated that MLB teams have in the past placed too low of a value on the *Intangibles* aspect of chemistry than the value of a win would suggest is appropriate.

Our efforts in this paper were largely descriptive. In future work, we plan to extend our methodology in order to allow for prediction by modeling the dynamics of team chemistry, including the possibility that organizational culture has its own set of dynamics as well. One way to capture the latter would be to include a team’s field and front office staff in our model. Furthermore, our analysis largely leveraged the playing time of individual players to explain teammate interactions and their impact on team performance. As a consequence, what is still left to understand is how to separately isolate the effect of players on their teammates and team performance through their off-the-field interactions.

7 Appendix

7.1 Data

Our data comprise 26,156 player-season observations over the 1998-2016 period. Nearly all players who participated in an MLB game during the 1998-2016 seasons appear in our analysis. The only exceptions are players who appeared in a game but failed to record an at-bat or an out. *fWAR* data come from the online database at fangraphs.com, and lineup information was constructed using the game-by-game data maintained at chadwick-bureau.com. All additional player, team, and performance information come from the databases maintained by Sean Lahman at seanlahman.com. While the Lahman database allows us to observe performance data by team for players that change teams within a season, FanGraphs only publishes *fWAR* at the season level of observation. In these cases, we divide a player’s season *fWAR* proportionally by his appearances for his respective teams, following the appearance weighting described in the main text. Thus, our dataset includes multiple observations within seasons for such players corresponding to each team on which they appear.
7.1.1 Adjacency Matrix

We utilize the weighted appearance measure from section 2.3 to construct the adjacency matrix that formalizes the network structure between teammates in our model. For a pairing between player $i$ and $j$ playing for team $n$ in season $t$, connection weights are defined as follows:

$$\alpha_{ij} = \sum_{n_i=n_j}^t (\kappa_i + \kappa_j)$$

$$\kappa_i = \begin{cases} l_i \ast PA_i + d_i \ast DOuts_i & \text{if } i \text{ is a position player} \\ POuts_i & \text{if } i \text{ is a pitcher.} \end{cases}$$

As noted in the text, a typical adjacency matrix is a symmetric matrix with 0’s on the diagonal and nonnegative values off the diagonal that denote connections between teammates. Instead, this expression states that the connection between two players is defined by not only their connection in the season in question, but also the sum of this value for each previous season in which the players were teammates. Whereas a value for a single season is typically 1, we allow for a richer definition of teammate connections by summing each pairing’s respective weighted appearances for the season in question. This captures two features of teammate connections. First, the more one or both players in a pairing play, the more likely they will have played together and the stronger their on-field connection will be. Second, the implicit orderings of our lineup ($l_i$) and defensive ($d_i$) weights shown in table 1 in the text capture specific on-field dynamics. If both players in a pairing tend to bat higher in the lineup, they will be more likely to affect each other’s performance. Similarly, pairings that include a catcher will be given relatively more weight, and if the other player is, for example, a middle infielder, this pairing will receive greater weight, all else equal, than one with a left fielder. Because pitchers receive a defensive weight equal to one, pitcher-catcher relationships will receive more weight than other position pairings, all else equal.

7.1.2 Regression Covariates

The regression analysis presented in sections 4.1 and 4.3 uses several covariates from the dataset that we construct from FanGraphs and the Lahman database. Our position indicators correspond to the position that the Lahman database indicates as the primary position for each player. Age is simply defined as the difference between the season year and the player’s birth year. Team and league indicators are pulled directly from the Lahman database, while we generate running totals for a players’ appearances in MLB and with their current team to control for experience and team tenure. Finally, manager indicators correspond to each team’s manager on opening day, thus ignoring managerial changes during the season. Tables 4 and 5
show the additional coefficients that correspond to the marginal effects shown in figure 8. The first column of table 5 shows the main effect for positional dummies, while the subsequent columns show coefficients for the interaction between the position dummies and the age polynomial.

### 7.2 A Spatial Factor Model

In matrix form, a spatial factor model can be written as

$$ Y = WFA + W\varepsilon $$

(21)

where $Y$ is an $ST \times N$ matrix of outcomes, $W$ is an $ST \times ST$ matrix of spatiotemporal weights, $F$ is an $ST \times K$ matrix of common factors, $\Lambda$ is an $K \times N$ matrix of factor loadings, and $\varepsilon$ is an $ST \times N$ matrix of idiosyncratic determinants of $Y$.

#### 7.2.1 The Reduced Form of a Spatial Autoregression

Equation 1 can be viewed as the reduced form of a spatial autoregression, or SAR. To see this, consider the following representation of a SAR

$$ Y = \rho AY + \nu $$

(22)

where $Y$ is a $ST \times N$ matrix of outcomes, $A$ is a $ST \times ST$ adjacency matrix, $\rho$ is a scalar parameter, and $\nu$ is an $ST \times ST$ matrix of residuals. Re-arranging the elements of equation 2, it can be rewritten

$$ Y = (I - \rho A)^{-1}\nu. $$

Defining $W \equiv (I - \rho A)^{-1}$ and assuming the approximate common factor structure $\nu = FA + \varepsilon$, equation 2 is shown to be equivalent to equation 1.

#### 7.2.2 Estimation

Estimation of equation 1 proceeds in two stages. In the first stage, we obtain an estimate of $\rho$ by maximum likelihood after imposing that the rows of the adjacency matrix $A$ sum to 1 and restricting its value to fall between -1 and 1. Combined, these normalizations satisfy the sufficient condition for $W$ to ensure that $(I - \rho A)$ be strictly diagonally dominant, i.e. $|1 - \rho A_{ii}| \geq \sum_{j \neq i} |-\rho A_{ij}|$. Given our estimate of $\rho$, we then proceed with spatial principal components analysis in the second stage assuming two common factors and
appropriate scale and sign normalizations on $\Lambda$. For the latter, we scale the factor loadings such that $\Lambda'\Lambda = I$; while for the former, we restrict the non-unit vector columns of $\Lambda$ to sum to zero.

Factor loading restrictions are handled in estimation by the expectation-maximization (EM) algorithm developed in [Dempster et al. (1977), Shumway and Stoffer (1982), and Watson and Engle (1983)] extended to include factor loading restrictions by [Reis and Watson (2010)]. To get a sense of how the EM algorithm operates in our case, consider the following: If the factors were known, then the factor loadings could be consistently estimated by a weighted least squares (WLS) regression of the form

$$\hat{\Lambda} = (F'W'WF)^{-1}(F'W'Y).$$  \hspace{1cm} (23)

Similarly, if the factor loadings were known, the factors are consistently estimated by

$$\hat{F} = (W^{-1}Y\Lambda')(\Lambda'\Lambda)^{-1}. \hspace{1cm} (24)$$

Given an unrestricted initial estimate of $\Lambda$ or $F$ and scale normalization, the EM algorithm iterates between these two WLS regressions until the sum of squared errors for equation 1 is minimized, imposing the factor loading restrictions at each iteration.

While the approximate factor structure we assume here is necessary for the EM algorithm to run, we can still use it to obtain the exact factor structure of our model by setting a convergence criterion which brings the sum of squared errors arbitrarily close to zero for a given number of common factors. This is achieved quite easily with our two factor model using a criterion which stops the algorithm when successive differences in the sum of squared errors are less than $1e^{-6}$.

References


Robinson, J. (1933), The Economics of Imperfect Competition, London: Macmillan.


Willis, A. (2017), ‘Passing the chemistry test’. http://www.slate.com/articles/sports/sports_nut/2017/05/team_chemistry_is_hard_to_quantify_when_will_sports_teams_figure_it_out.html.
<table>
<thead>
<tr>
<th></th>
<th>Team Wins Regression 1998-2016</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Team $fWAR$</td>
<td>0.997 (0.021)</td>
<td>Team $fWAR^-$</td>
<td>1.031 (0.015)</td>
</tr>
<tr>
<td>Constant</td>
<td>47.682 (0.733)</td>
<td>48.151 (0.517)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.799 0.887</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>570 570</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses
<table>
<thead>
<tr>
<th>Position</th>
<th>d value</th>
<th>Lineup Slot</th>
<th>l value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catcher</td>
<td>0.214</td>
<td>1</td>
<td>0.212</td>
</tr>
<tr>
<td>First Base</td>
<td>0.036</td>
<td>2</td>
<td>0.187</td>
</tr>
<tr>
<td>Second Base</td>
<td>0.143</td>
<td>3</td>
<td>0.162</td>
</tr>
<tr>
<td>Third Base</td>
<td>0.143</td>
<td>4</td>
<td>0.136</td>
</tr>
<tr>
<td>Short Stop</td>
<td>0.179</td>
<td>5</td>
<td>0.111</td>
</tr>
<tr>
<td>Left Field</td>
<td>0.071</td>
<td>6</td>
<td>0.086</td>
</tr>
<tr>
<td>Center Field</td>
<td>0.143</td>
<td>7</td>
<td>0.061</td>
</tr>
<tr>
<td>Right Field</td>
<td>0.071</td>
<td>8</td>
<td>0.035</td>
</tr>
<tr>
<td>Designated Hitter</td>
<td>0</td>
<td>9</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table 3: Player Salary Regressions

<table>
<thead>
<tr>
<th></th>
<th>(1) log(salary)</th>
<th>(2) log(salary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( FA_0 \times \text{Career } fWAR )</td>
<td>0.09***</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>( FA_1 \times \text{Career } fWAR )</td>
<td>0.12***</td>
<td>0.12***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>( FA_2 \times \text{Career } fWAR )</td>
<td>0.04***</td>
<td>0.02***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>( FA_0 \times \text{Career } pcWAR )</td>
<td>-0.43***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td>( FA_1 \times \text{Career } pcWAR )</td>
<td>-0.39***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>( FA_2 \times \text{Career } pcWAR )</td>
<td>0.09*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>( FA_0 \times \text{Career } (pcWAR - \xi) )</td>
<td>-0.52**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td></td>
</tr>
<tr>
<td>( FA_1 \times \text{Career } (pcWAR - \xi) )</td>
<td>-0.41***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>( FA_2 \times \text{Career } (pcWAR - \xi) )</td>
<td>0.39***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>( FA_0 \times \text{Career } \xi )</td>
<td>-0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td>( FA_1 \times \text{Career } \xi )</td>
<td>-0.38***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>( FA_2 \times \text{Career } \xi )</td>
<td>-0.23***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>17,170</td>
<td>17,170</td>
</tr>
<tr>
<td>Within-group ( R^2 )</td>
<td>0.73</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Select variable estimates reported. Specifications include contract status indicators, indicators for years with current team, and defensive position indicators interacted with age, games of MLB experience, and experience squared. Robust standard errors clustered on player in parentheses.

*** p<0.01, ** p<0.05, * p<0.1
Table 4: Select $pcWAR$ Regression Estimates

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$pcWAR$</td>
</tr>
<tr>
<td>$fWAR$</td>
<td>0.0933***</td>
</tr>
<tr>
<td></td>
<td>(0.000717)</td>
</tr>
<tr>
<td>MLB Experience</td>
<td>-7.44e-05***</td>
</tr>
<tr>
<td></td>
<td>(3.86e-06)</td>
</tr>
<tr>
<td>Team Experience</td>
<td>-6.97e-05***</td>
</tr>
<tr>
<td></td>
<td>(6.95e-06)</td>
</tr>
<tr>
<td>Observations</td>
<td>26,156</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.773</td>
</tr>
</tbody>
</table>

Team, league, and manager fixed effects included. Robust standard errors clustered on player in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Table 5: Age-position interaction coefficients

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Age²</th>
<th>Age³</th>
<th>Age⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B</td>
<td>-2.098*</td>
<td>0.277***</td>
<td>-0.0135**</td>
<td>0.000291**</td>
</tr>
<tr>
<td></td>
<td>(1.108)</td>
<td>(0.137)</td>
<td>(0.00625)</td>
<td>(0.000124)</td>
</tr>
<tr>
<td>2B</td>
<td>-0.402</td>
<td>0.0899</td>
<td>-0.00592</td>
<td>0.000151</td>
</tr>
<tr>
<td></td>
<td>(1.775)</td>
<td>(0.237)</td>
<td>(0.0117)</td>
<td>(0.000252)</td>
</tr>
<tr>
<td>3B</td>
<td>-1.438</td>
<td>0.208</td>
<td>-0.0108</td>
<td>0.000240</td>
</tr>
<tr>
<td></td>
<td>(3.223)</td>
<td>(0.456)</td>
<td>(0.0239)</td>
<td>(0.000551)</td>
</tr>
<tr>
<td>C</td>
<td>2.752</td>
<td>-0.342</td>
<td>0.0159</td>
<td>-0.000330</td>
</tr>
<tr>
<td></td>
<td>(2.133)</td>
<td>(0.289)</td>
<td>(0.0145)</td>
<td>(0.000321)</td>
</tr>
<tr>
<td>CF</td>
<td>-0.339</td>
<td>0.033</td>
<td>-0.000753</td>
<td>-1.06e-05</td>
</tr>
<tr>
<td></td>
<td>(3.105)</td>
<td>(0.426)</td>
<td>(0.0216)</td>
<td>(0.000479)</td>
</tr>
<tr>
<td>DH</td>
<td>0.998</td>
<td>0.171</td>
<td>-0.00987</td>
<td>0.000238</td>
</tr>
<tr>
<td></td>
<td>(2.807)</td>
<td>(0.372)</td>
<td>(0.0182)</td>
<td>(0.000391)</td>
</tr>
<tr>
<td>LF</td>
<td>-2.902</td>
<td>0.380</td>
<td>-0.0182</td>
<td>0.000376</td>
</tr>
<tr>
<td></td>
<td>(2.645)</td>
<td>(0.347)</td>
<td>(0.0168)</td>
<td>(0.000353)</td>
</tr>
<tr>
<td>RF</td>
<td>-0.428</td>
<td>0.052</td>
<td>-0.00193</td>
<td>2.26e-05</td>
</tr>
<tr>
<td></td>
<td>(2.265)</td>
<td>(0.308)</td>
<td>(0.0156)</td>
<td>(0.000346)</td>
</tr>
<tr>
<td>RP</td>
<td>1.667***</td>
<td>-0.202***</td>
<td>0.00917***</td>
<td>-0.000182***</td>
</tr>
<tr>
<td></td>
<td>(0.515)</td>
<td>(0.0674)</td>
<td>(0.00327)</td>
<td>(6.98e-05)</td>
</tr>
<tr>
<td>SP</td>
<td>1.305*</td>
<td>-0.149</td>
<td>0.00629</td>
<td>-0.000119</td>
</tr>
<tr>
<td></td>
<td>(0.711)</td>
<td>(0.0911)</td>
<td>(0.00428)</td>
<td>(8.73e-05)</td>
</tr>
<tr>
<td>SS</td>
<td>5.785</td>
<td>-0.788</td>
<td>0.0397</td>
<td>-0.000888</td>
</tr>
<tr>
<td></td>
<td>(4.365)</td>
<td>(0.611)</td>
<td>(0.0316)</td>
<td>(0.000716)</td>
</tr>
<tr>
<td>Utility</td>
<td>4.251</td>
<td>-0.581</td>
<td>0.0298</td>
<td>-0.000673</td>
</tr>
<tr>
<td></td>
<td>(2.907)</td>
<td>(0.403)</td>
<td>(0.0206)</td>
<td>(0.000458)</td>
</tr>
</tbody>
</table>
Figure 1: Complementarity of Select Player Relationships
Figure 2: Team Productivity Residuals
Ranked on a scale from 0 to 100, with 100 equal to best organization over 1998-2016 seasons

Figure 3: MLB Team Chemistry Factor Loadings
Figure 4: Team Chemistry and Wins-above-Replacement
Solid red lines are 45 degree lines. Vertical lines denote thresholds for Scrub/Role (fWAR=1) and Good/Star (fWAR=4) players.

Figure 5: $fWAR^-$ and $fWAR^+$ vs. $fWAR$
fWAR+ = fWAR + pcWAR

Vertical lines denote fWAR thresholds for Scrub/Role (fWAR=1) and Good/Star (fWAR=4) players.

Figure 6: Player Chemistry and Wins-above-Replacement
Figure 7: pcWAR Rankings
Average Marginal Effects with 95% CIs

Conditional on fWAR, MLB and team games played, league, manager and team

Figure 8: Age-Position Team Chemistry Profiles
Intangibles are the residuals from pcWAR regressed on fWAR, MLB and team games played, age-position profile, league, manager and team.

Figure 9: Intangibles Rankings
Intangibles are the residuals from pcWAR regressed on fWAR, MLB and team games played, age-position profile, league, manager and team.

Figure 10: David Ross’ Chemistry Profile