

TI 2015-115/VII
Tinbergen Institute Discussion Paper

Matching and Winning? The Impact of Upper and Middle Managers on Team Performance in Major League Baseball

Revision: March 2020

Thomas Peeters¹

Steven Salaga²

Matthew Juravich³

¹ Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute, the Netherlands;

² Texas A&M University, United States

³ University of Akron, United States

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Matching and Winning? The Impact of Upper and Middle Managers on Firm Performance in Major League Baseball

We investigate the joint impact of managers at different hierarchical levels on firm performance in Major League Baseball. We separately quantify the contribution of upper and middle managers and the impact of their “match quality” – the degree to which managers cooperate effectively across layers to impact firm success. We establish that match quality is a statistically significant and economically meaningful driver of firm performance. Higher quality managers tend to be matched together across levels and achieve higher match quality during their joint employment. Match quality does not improve over the length of a joint employment spell, but lower match quality is found in pairs with more divergent educational attainment and prior strategic approaches. Hence, match quality is partly innate and manager pairings may have difficulty improving their cooperation through learning. When we control for match quality, we find significantly lower estimates of heterogeneity in manager ability compared to commonly used estimators of managerial impact. Still, both middle and upper managers retain a meaningful impact on firm performance.

Keywords: match quality, upper management, middle management, firm performance

Acknowledgements: The authors thank Department Editor Bruno Cassiman, the Associate Editor and three anonymous referees, as well as Oriana Bandiera, Alan Benson, Christophe Boone, Jan Bouckaert, Charles Brown, Tine Buyl, Josse Delfgaauw, Amanda Goodall, Brian Mills, Andrew Oswald, Stefan Szymanski, Jason Winfree and Nick Zubanov for interesting discussions and comments on this project. Further feedback from seminar/conference participants at Tilburg University, Erasmus School of Economics, the University of Antwerp, Tinbergen Institute, the 2015 Eastern Economic Association, 2015 NASSM, 2015 SOLE/EALE, and 2015 EASE meetings was also greatly appreciated. Peeters gratefully acknowledges financial support from the Flanders Research Foundation during part of this research project.

1. Introduction

How large is the influence of managers at different organizational levels on firm performance? How much does the cooperation between managers across hierarchical levels impact firm success? Do high-quality managers work together and do they cooperate effectively? Is high-quality cooperation in a management pairing innate in nature or can managers improve the quality of their cooperation over time through learning? To examine these questions, we quantify the individual contribution of middle and upper managers to firm success while concurrently controlling for the quality of the cooperation between them.

The strategic management literature has paid considerable attention to measuring how much managers at different levels of the organizational hierarchy impact firm policies and performance (e.g., Hambrick, 2007; Wooldridge et al., 2008; Wooldridge & Floyd, 1990). Extant research has examined the influence of upper managers, such as Chief Executive Officers (CEOs) and Chief Financial Officers (CFOs) (Hambrick & Mason, 1984; Bertrand & Schoar, 2003), and middle managers such as team leaders (Crocker and Eckardt, 2014; Lazear et al., 2015), branch managers (Siebert & Zubanov, 2010; Owan et al., 2014) and operations managers (Aime et al., 2010; Hendricks et al., 2014). Relatedly, observable characteristics such as education (Chevalier & Ellison, 1999; Mair, 2005; Goldfarb & Xiao, 2011; Juravich et al., 2017), work experience (Goodall et al., 2011; Goodall & Pogrebna, 2014), reputation (Falato et al., 2015) and personality traits (Malmendier et al., 2011; Kaplan et al., 2012), have all been linked to managerial performance at different hierarchical levels. To date, this literature has largely examined the impact of various layers of management in isolation despite the reality that managers along the firm hierarchy jointly influence the performance of the organization (Hitt et al., 2007). The current study therefore attempts to identify the joint impact of managers across organizational levels and to measure the importance of their cooperation. Thus, this focus addresses an understudied topic in this literature.

As a foundation for our theory, we consider the skills, abilities, and experiences of individual managers as human capital (Ployhart & Moliterno, 2011). When analyzing multiple levels of management, the combination of individual human capital at the group level can serve as a source of differentiation (Wright & McMahan, 2011). We propose that group human capital, the combination of individual-level attributes for group-level action (Ployhart et al., 2014), can lead to a competitive advantage for those firms where managerial alignment is optimized. As such, human capital resource combinations across managerial layers are argued to directly impact firm level outcomes via “match quality” - the degree to which the

combination of specific attributes possessed by individual managers leads to effective cooperation. Thus, we posit that the match quality between managerial levels shapes firm performance via the extent to which managerial pairs cooperate and ultimately, make decisions effectively.

Given that the empirical importance of match quality has been demonstrated in a variety of related contexts, such as matching between CEOs and firms (Pan, 2017; Terviö, 2008), manager risk preference and firm compensation policies (Bandiera et al., 2015), and supervisors and production workers (Crocker & Eckardt, 2014; Lazear et al., 2015), a natural extension is to analyze the interaction between managers along the firm hierarchy. By investigating the impact of match quality between middle and upper managers, we expand the scope of the extant strategic management literature while concurrently contributing to our understanding of how human capital resources can be leveraged to enhance firm performance (Ployhart et al., 2014).

Our empirical analysis uses data from Major League Baseball (MLB) to separately quantify the impact of the middle and upper manager and the match quality between them on firm performance. We find that match quality is an economically meaningful and statistically significant driver of performance heterogeneity. Improving the match quality in a manager pairing by one standard deviation increases firm performance as much as moving up one standard deviation in the quality of either manager. Taking match quality into account reduces the estimated heterogeneity in manager quality by about 60% for middle managers and 25% for upper managers. This finding emphasizes the importance of integrating the group human capital perspective into the analysis of managerial impact.

Match quality does not increase along the duration of a joint employment spell, implying that “learning” is ineffective in advancing the cooperation of a manager pairing. The qualities of the middle and upper manager in a pairing are positively correlated with one another and with firm quality. Match quality itself is also positively correlated to the qualities of both managers. In other words, better managers tend to work together and achieve higher match quality in their cooperation. As such, individual and group level human capital appear to be complements and not substitutes. Manager pairs with large discrepancies in educational attainment and divergent propensities to use specific game strategies attain lower match quality. This is consistent with the view that the estimated match effects serve as a proxy for productive cooperation across organizational levels, which in turn contributes to firm performance. Finally, we show that match quality translates into labor market outcomes, as spells with lower match quality tend to be shorter lived.

This paper proceeds as follows. First, we discuss our conceptualization of manager and match quality and outline how managerial cooperation across levels may impact firm performance. We then discuss the role of managers in professional baseball organizations and introduce a model which serves as the basis for our empirical analyses. Next, we detail our empirical implementation and explain our results. Finally, we summarize our findings, discuss how the results are applicable to firms in a variety of industries, and offer concluding thoughts.

2. Match Quality and Firm Performance

Since match quality is derived from the cooperation of managers across hierarchical levels, we first need to define “manager quality” at the individual level. In a broad sense, the term accounts for the talent or ability level of the individual manager. In line with upper echelons theory, ability results from the unique personal experiences of each manager including factors such as education, functional and technical experience (Hambrick & Mason, 1984). Individual ability directly translates into a manager’s capability to impact firm outcomes with higher quality managers having greater ability to positively influence firm performance. This interpretation inevitably leads to the question whether manager quality is constant over the course of a manager’s career, as learning-on-the-job could improve a manager’s capability to impact the firm over time (Chatterji et al., 2016). Alternatively, manager heterogeneity could represent individual management “styles”, which reflect the idiosyncrasies of individual managers along various dimensions (Bertrand & Schoar, 2003). In our conceptualization, these idiosyncrasies could contribute to manager quality (e.g., when they are understood as talent or ability) as well as match quality (e.g., when they refer to how manager personalities or policy preferences align within the firm).

Match quality is a consequence of the cooperation between managers at different hierarchical levels and its interplay with the firm. In this capacity, it represents both the degree to which the interaction across managerial levels and the alignment between both managers and the firm impacts performance. The quality of a match is partially determined by the sorting process of dynamic labor markets as some managers are under contract as new matches are created. Still, once managers have matched across levels, their cooperation may generate a level of success that has the ability to either positively or negatively impact firm outcomes (Raes et al., 2011). If match quality varies according to the same factors which influence manager quality, the degree of complementarity between managers across levels could determine how well they are able to convert their separate abilities into actions which positively impact the firm. Likewise, upper and middle managers with conflicting visions on which managerial approach to

implement may find it harder to cooperate effectively towards improving firm performance. As such, we propose that the interaction of manager qualities and the extent to which they are appropriately “matched” in terms of personal background and strategic approach may serve as a human capital resource (Ployhart et al., 2014).

The fact that our definition of match quality captures the effect of management cooperation across hierarchical levels sets our analysis apart from existing research on worker complementarity. At the horizontal co-worker level, Herbst and Mas (2015) summarize the literature examining spillover effects and the positive returns of worker productivity to peer production. However, these workers are performing production-level tasks, meaning the need to actively align human capital resources with firm objectives is largely absent. There are numerous ways in which managerial cooperation at different levels, as examined here, is potentially more complex in nature and may have a different impact on firm performance. Across hierarchical levels, Lazear et al. (2015) show the positive effects of high-quality supervisors on both team productivity and employee retention. Similarly, Artz, et al. (2017) demonstrate positive effects of supervisor competence on the satisfaction of direct-report workers. Crocker and Eckardt (2014) illustrate positive co-worker effects both within and across levels, as they establish that worker performance is partially dependent on co-worker skill as well as the quality of the direct supervisor. Our work is unique in comparison to these studies, in that we examine cooperation across two organizational levels, which are both pivotal in selecting and developing human capital resources in alignment with the firm’s objectives.

Thus, this study contributes to the literature by demonstrating that cooperative interaction across hierarchical levels of management is economically significant and impacts firm performance above and beyond individual manager effects. We also demonstrate that estimates of manager effects may be overstated when not accounting for the match quality between different managerial layers and the organization. As such, we contribute to a recent call for scholars to study human capital resource combinations (Ployhart et al., 2014).

3. Model

3.1. The Role of Upper and Middle Managers in Major League Baseball

In our empirical analysis, we use a dataset on MLB managers. MLB franchises operate under a relatively uniform organizational hierarchy, which has been stable over the past several decades.¹ At the top of the firm, an owner or ownership group establishes a formal hierarchy of employees to serve as a chain of command for decision making. Ownership typically appoints an upper level executive – the general manager (GM) – who directs long-term strategy. GMs also serve as upper-level managers with respect to personnel management. They oversee and facilitate the hiring and firing of employees (both other managers and players) based upon the alignment of employee abilities with firm goals. A clearly defined middle manager, called the coach or manager², is responsible for the day-to-day management of the team. Coaches are granted autonomy in organizing their human resources to pursue competitive success in the most effective manner possible.

MLB managers perform largely similar tasks (e.g., planning, organization, and recruitment) and take on comparable roles (e.g., communicator, motivator, etc.) to executives in many other industries. For example, MLB managers at both levels hire a support staff. The GM oversees personnel specialized in baseball operations, player development, scouting and analytics. The coach employs a staff of assistants including a bench coach and hitting and pitching specialists. The organizational goals of MLB teams mirror those of non-sport organizations as they pursue financial success by optimizing efficiency in their competitive environment. As such, MLB teams are engaged in a continuous cycle of performance improvement. One common approach in this respect is to change the team’s strategic direction via the replacement of the coaching staff. The GM should then compile a roster of appropriately skilled players to support the approach of the new coach. An optimal roster would be populated with players whose attributes complement the (game) strategies jointly identified by the middle and upper manager. As such, both GMs and coaches occupy key managerial roles which directly impact strategy formulation and implementation, financial investment, and human resource management. Thus, there is a clear need for cooperation in managerial decision making across organizational levels. Hence, we propose that match quality between managers may be empirically relevant in this setting.

¹ This is in contrast to many other industries, as documented in Guadalupe et al. (2014).

² In baseball the head coach is often called the “manager.” To avoid confusion between the manager and general manager, we refer to the “manager” as the “coach”. When simultaneously referring to both management levels, we use the term “managers.”

MLB is particularly appealing as a research context because the industry generates an abundance of data on (a) firm inputs (e.g., player wages), (b) firm outputs (e.g., scores, results) and (c) managers' personal characteristics.³ The structural homogeneity of MLB further ensures that managers are rehired in the same role across different firms. Combined with frequent manager turnover, this results in numerous co-worker matches which facilitates the separate identification of manager and match effects. Finally, the aforementioned similarities between MLB teams and firms in other industries broaden the applicability of our results beyond the scope of this context.

3.2. An Empirical Model of Firm Production in Major League Baseball

We specify an empirical model for the output production of MLB teams, which we define as the on-field performance (y_{git}) in a game g played between team i and j during season t . We control for five inputs; labor (L_{git}), capital (K_{it}), home advantage (γ_{gi}), fixed firm-specific factors (ω_i) and management. We explicitly separate management from labor inputs, such that labor only refers to the workers (players) the team employs. We allow labor inputs to vary game-by-game, depending on the line-up, but assume capital is fixed over the season. We include home advantage, because it is prevalent across a wide range of sports and remarkably stable over time (Pollard and Pollard, 2005). Fixed firm-specific factors include elements such as team history, fan loyalty and organizational culture. Management inputs encompass the contribution of the coach at the middle level (μ_i) and the GM at the upper level (v_i) of the firm. We finally control for the match quality (σ_i)⁴ between the team, its current coach and current GM, which may impact the firm above and beyond the simple addition of manager abilities.

Following Peeters and Szymanski (2014), we assume that input use only matters in relative terms. In other words, a team's inputs only impact the game's result in relation to the inputs of the opposing team. If a factor equally affects both teams (e.g., game day weather), our specification washes away its absolute effect. Our model thus reduces to estimating the fraction of two production functions,

$$\tilde{y}_{git} = L_{git}^{\beta_l} K_{it}^{\beta_k} \exp(\gamma_{gi} \omega_i \mu_i v_i \sigma_i), \quad (1)$$

where the return parameters β_l and β_k measure the impact of labor and capital.

³ See Kahn (2000) for a more elaborate discussion on the use of sports data in managerial/labor economics. The literature has studied a variety of questions using sports data, e.g. the durability of competitive advantage (Aime et al., 2010) and human resource practices (Staw & Hoang, 1995).

⁴ To simplify notation we suppress the indices for the coach and GM in the match quality term.

We first rewrite equation (1) in logs and take the difference between the resulting functions of both teams in the game. We then add a game-specific error term (ε_{gijt}), which captures the chance factors inherent in sports competition. This yields our baseline estimation model with additive terms for inputs, managers, teams and match quality,

$$y_{gijt} = \gamma_{gi} - \gamma_{gj} + \beta_l(l_{git} - l_{gjt}) + \beta_k(k_{it} - k_{jt}) + \omega_i - \omega_j + \mu_i - \mu_j + v_i - v_j + \sigma_i - \sigma_j + \varepsilon_{gijt}. \quad (2)$$

4. Empirical implementation

4.1. Identifying the Impact of Managers and Match Quality

A common approach to identify the contribution of firms and workers to firm-level outcomes is to include firm and worker fixed effects in a linear model such as equation (2) (e.g., Bertrand & Schoar, 2003). As we are interested in heterogeneity at two separate organizational levels, we introduce three sets of dummy variables to estimate the individual fixed effect of each team, coach and GM. It is not feasible to estimate yet another set of dummy variables for each coach-GM-team match, as these would be perfectly collinear with the manager and firm effects. In this approach, match quality can therefore only be gauged through the average residual of a manager pairing.

As shown by Abowd et al. (1999), the identification of worker and firm effects hinges on observing mobile workers in different combinations of employers and co-workers. In our case, we require mobility in two dimensions. First, each firm has to be paired with at least one manager who is also employed by another firm. We therefore search the data for networks of firms connected through moving managers. For each firm in such a network, we can separate the firm and manager effects. Second, each GM has to be paired with at least one coach who has worked with a different GM, and each coach has to be paired with at least one GM who worked with a different coach. We therefore look for networks of GMs connected by moving coaches, and networks of coaches connected through moving GMs. By construction, firms and managers in such a connected network share a common reference to which all person and firm effects are scaled, while this is not the case across different networks. In our application, a natural choice for the reference category are the so called “caretaker” or interim managers, who did not secure a long-

term management position in MLB. As we expect these managers to have a minor impact on firm performance, they constitute a natural lower bound for the impact of managers with established careers.⁵

Our sample covers all games played in MLB between 1988 and 2012 for a total of 148 coaches, 114 GMs⁶ and 31 teams combined into 345 employment spells. Slightly under 38% of coaches and 24% of GMs are ‘movers’, who worked for two or more teams. We drop the three teams, which lack a mover at one of their management levels.⁷ All remaining clubs are connected in the same network via moving coaches, whereas for GMs, there are two separate networks.⁸ At these teams, 146 coaches and 105 GMs have been paired with a manager at the other level, which has at least one other co-worker in the data. We exclude all coach-GM combinations who fail to meet this requirement, which leaves 136 coaches (92% of the original total) and 94 GMs (82% of the total) in the final dataset. Each game appears twice in the data, once from each team’s perspective. The entire procedure shrinks the number of games in the sample from 55,106 (or 110,212 observations) down to 42,302 (84,604 observations). Finally, we sort approximately 21% of coaches (with less than 130 observations) and 15% of GMs (with less than 260 observations) into the caretaker reference category. These represent 1,995 and 1,365 game observations respectively.

A downside of using fixed effects to estimate equation (2) is that match quality must be absorbed by the error term. Hence, the mean match quality for a given manager has to equal zero across all the manager’s pairings. In a hypothetical dataset, where each manager is observed in many different pairings, this assumption need not be problematic, as good and bad matches may cancel out in the long run. In a typical dataset however, each manager is observed in only a handful of pairings. Thus, the fixed effects approach runs the risk of attributing a considerable portion of the match value to manager and firm effects. As a result, the importance of manager and firm effects may be overstated if these are positively correlated with match quality.

To overcome this issue, we use the two step “random effects” (RE) estimator developed by Jackson (2013). In the first step, we introduce a composite spell fixed effect (φ_i), which functions as an indicator

⁵ Usually caretaker managers fail to get permanent positions as a result of poor team performance under their reign. A further advantage of choosing caretakers as the reference, is that we do not have to estimate their individual effects from the limited amount of observations available per individual caretaker, which reduces the risk of incidental parameter bias.

⁶ One GM observation is in fact a duo, i.e., Mike Flanagan and Jim Beattie of the Baltimore Orioles in the early 2000’s.

⁷ These are Minnesota, Pittsburgh and Tampa Bay.

⁸ GM network 1 is Texas, Milwaukee and Cleveland, while network 2 is all other MLB clubs. To arrive at comparable estimates, we repeat the fixed effect procedure using each possible pairing of firms across both GM networks as the reference firm and report the average effects across all estimates. The correlations across fixed effects estimates with different reference firms is always above 0.95, so, empirically, this appears to be a fairly minor issue.

for each unique pairing between a firm, middle, and upper manager. This spell fixed effect replaces all firm and manager effects in equation (2), such that it reduces to,

$$y_{gijt} = \gamma_{gi} - \gamma_{gj} + \beta_l(l_{git} - l_{gjt}) + \beta_k(k_{it} - k_{jt}) + \varphi_i - \varphi_j + \tilde{\varepsilon}_{gijt}. \quad (3)$$

We estimate this model by OLS and isolate both spell fixed effects (φ_i and φ_j) and the error term ($\tilde{\varepsilon}_{gijt}$). We sum these three terms at the level of each observation, which creates the dependent variable for the second step (η_{gijt}). We then decompose this joint spell-error term into the coach, GM, team and match effects for team i and j in a mixed effects model estimated by maximum likelihood. We allow for time variation in match quality over the employment spell by including a third-degree polynomial (h) of spell duration for both teams (d_{gi} and d_{gj}). The second stage model we estimate is then given by

$$\eta_{gijt} = \omega_i - \omega_j + \mu_i - \mu_j + v_i - v_j + \sigma_i - \sigma_j + h(d_{gi}) - h(d_{gj}) + \varepsilon_{gijt}, \quad (4)$$

$$\text{with: } \eta_{gijt} = \varphi_i - \varphi_j + \tilde{\varepsilon}_{gijt}.$$

This procedure uses the observed variation in manager, firm and match qualities to form an estimate of the match effects for managers with a limited number of spells. This replaces the unrealistic assumption that the mean match quality for each manager equals zero. While this “random effects” procedure changes the way we attribute variation in the data across different factors, it does not introduce a fundamentally different identification strategy. The identification of individual effects still depends on observing moving managers as described above. We therefore apply both procedures (random and fixed effects) to the same data sample. This facilitates the interpretation of our results, as we can compare estimates across approaches as well as contrast our estimates against previous work using fixed effects.

4.2. Measuring Outputs and Inputs

Table 1 shows summary statistics for our estimation sample. In baseball, at the end of the game, the team which has scored the most ‘runs’, wins. If both teams have scored an equal number of runs, the game is prolonged until a winner can be declared. Our first output measure, ‘run difference’, is the difference between the runs scored by both teams. A negative value indicates a loss and a positive value implies a win. Our dataset contains one observation from the perspective of each team in a given game, so every observation of run difference has a diametrically opposed observation for the opposing team. By design, the average run difference is equal to zero, and the maximum and minimum are perfectly symmetric around zero. The game-level standard deviation is slightly above four. A second measure of team production is the ‘game result’, which takes a value of zero for a loss and one for a win. Again by design,

the mean and standard deviation are both equal to 0.5.⁹ MLB teams play one another multiple times per season. As these games are not completely independent events, we use clustered standard errors for all matches between the same two teams in a particular season.

<Insert Table 1 here>

We follow the economic literature on sports by measuring labor inputs through player salaries. We expect a positive effect of a team's total wage bill on performance (Szymanski, 2003).¹⁰ The average team payroll in the sample stands at approximately \$75m per year, but some large market teams (e.g. New York Yankees) were spending over \$200m by the end of the sample period. Following Bloom (1999) and Depken (2000), we expect that a high intra-team standard deviation of player wages is associated with lower on-field performance. The average intra-team standard deviation is fairly modest at \$3m with a standard deviation of around \$1.5m. A specific feature in baseball is the role of the most important defensive player, the starting pitcher. As fatigue diminishes a pitcher's performance, teams rotate their starting pitchers from game to game. The pitcher starting the game has a large impact on the game outcome, so we separately account for the salary of the starting pitcher. The average salary for these players is \$2.9m per season, but the heterogeneity is very wide because superstar pitchers command substantial wages. The main capital input of MLB teams is their stadium. We proxy the current value and opportunity cost of a stadium by its age and the average home value in the club's metropolitan statistical area (MSA). Both measures show significant variation with the oldest stadium being in use for over a century.

<Insert Table 2 around here>

Table 2 provides summary statistics for the personal characteristics, career characteristics and strategic approaches of the GMs and coaches in our estimation sample. GMs are more educated than coaches, as more have attended and graduated college. We also indicate whether a GM attended an institution with a strong academic reputation (i.e., a top 100 school in the 2014/15 QS world university rankings). Coaches have more technical (playing) experience, as more of them played professionally, played in MLB and appeared in the MLB All-Star Game.¹¹ Approximately 15% of all coaches have a minority background, while this number stands at only 7.9% for GMs. The average of career run difference and win

⁹ We present results for both measures throughout the paper. This assures our findings are robust to the choice of output measure and it aligns our results with common measures of team success, such as 'wins produced' or seasonal 'win percent'.

¹⁰ The MLB player market is not regulated by a salary cap, which results in large discrepancies in payroll across teams. MLB has a luxury tax, which imposes an additional tax on salary payments above a certain threshold. However, only two teams (NY Yankees and Boston Red Sox) regularly pay a contribution and the tax represents a very minor share of overall payroll costs.

¹¹ The All-Star Game fields the best players of the season across all teams against one another.

percentage is negative, because better performing managers tend to have longer careers, which drives down the average on a per-manager basis. The standard deviation of firm performance is comparable for coaches and GMs, at about 0.4 for run difference and 4%-5% for game results.¹² On average, GMs have more managerial experience than coaches, but a typical GM is slightly younger than his coaching counterpart. We finally characterize the strategic approach of managers using two offensive game statistics (the number of sacrifice hits and stolen base attempts) and two defensive statistics (errors and intentional walks).¹³

5. Empirical Analysis and Results

5.1. First Stage Results

Table 3 presents estimation results of the first stage model with run difference as the dependent variable. In columns (1)-(5), we sequentially add team, coach and GM fixed effects until all are contained in a single model. This represents the full fixed effects model specified in equation (2). The F-test results for the various levels of fixed effects indicate that both upper and middle managers significantly impact production. The sixth and final column includes the combined spell fixed effects, which capture each unique coach-GM-team combination. This is the first stage estimation of equation (3).

<Insert Table 3 around here>

Our estimates for the input coefficients largely correspond to the hypothesized effects. We find a significant positive effect for home advantage, total team payroll and starting pitcher salary. The coefficient on the standard deviation of total team payroll is negative and in most cases significant. We uncover a positive effect for home value in the team's MSA, but this is insignificant in several specifications. Finally, there is no consistent effect of stadium age. Switching to the win/loss dummy for the dependent variable does not alter the sign or significance of any input coefficient. We provide further details and robustness checks on these results in section 1 of the online appendix.

¹² In terms of personal background, the managers in the estimation sample closely resemble those in the full sample. However, their career performance is relatively less dispersed, because we drop several poor performing managers with short careers.

¹³ Sacrifice hits occur when a team gives up an out to advance a base runner. Attempts to steal bases (= stolen bases + caught stealing) involve a runner trying to advance a base without the ball being put into play. Errors represent defensive miscues, which may occur in a variety of manners. We consider errors a defensive strategy as a common approach in MLB is to neglect defensive ability in favor of offensive ability when hiring workers. A team issues an intentional walk when it allows the current hitter to advance to first base in order to pitch to the next hitter in the lineup. We selected these variables because (a) managers vary strongly in their application and (b) they are not obviously connected to game outcomes, as would be the case for variables which measure successful actions, e.g. stolen bases. See www.retrosheet.org for more information on these variables.

5.2. The Impact of Firms, Managers and Match Quality on Firm Performance

5.2.1. How Heterogeneous are Firms, Managers and Matches?

We gauge the heterogeneity among managers, firms and matches through the standard deviation of their estimated individual effects. Note however that a standard deviation can never take a negative value. Hence, even if firm performance would entirely be driven by random events, the standard deviation of the estimated firm, manager and match effects would still be larger than zero. It is therefore not very informative to compare the heterogeneity in the estimated effects to zero, because in real-world datasets the amount of observations is never sufficient to wash away the full impact of randomness. Instead, Fitza (2014) suggests to compare the individual heterogeneity estimated on real-world data to the heterogeneity estimated on placebo datasets, where the dependent variable is randomly drawn from the same distribution as the original performance variable. The placebo estimates then function as a comparison point indicating how large the estimated heterogeneity in the real-world estimates should be to rule out that they are entirely driven by randomness. For our application, we create 500 placebo datasets in which we randomly reassign game results across observations. We then compare the standard deviation of the estimated firm, manager and match effects from the actual data to the average standard deviation for the same model estimated in each of these 500 placebo datasets.

In Table 4 we report the results of this exercise for three versions of the random effects model, one which excludes both the match effects and spell duration polynomial, one which includes match quality, but excludes the duration polynomial and the full model, which includes both. For the fixed effects model, which ignores match quality, we test two specifications, one which excludes and one which includes the spell duration polynomial.

<Insert Table 4 around here>

First, we note the impact of match quality is itself substantial. For both outcome measures, the standard deviation of match effects is on par with that of GMs, larger than that of coaches and teams, and a magnitude larger than what is found in the placebo data. A standard deviation improvement in match quality is equivalent to an increase of approximately 1.9 wins per season (1.15% times 162 games). Given that the typical range between the best and worst performing team is around 35 to 40 wins over an entire season, match quality explains about 5% of the overall difference between teams. Recent estimates put the cost of one additional win per season at approximately \$5m in player salary (Cameron, 2014), so an improvement by 1.9 wins represents a value of \$9.3m or 8% of average team payroll in 2015. By comparison, hiring a manager who is one standard deviation more effective, increases the win percentage

of the team by 0.95% for middle managers and by 1.28% for upper managers in the fully specified model. This corresponds to a salary investment of \$7.6m and \$10m respectively. Hence, both managerial ability and the match quality between managers are economically relevant components of firm performance.

Accounting for match quality reduces the estimated heterogeneity in manager ability for both upper and middle managers. In the random effects model, adding controls for match quality reduces the estimated standard deviation of middle manager ability by about 60%, while the reduction for upper managers is approximately 25%. This drop is larger than any movement found in the placebo data, and therefore cannot be attributed to randomness. If we compare the findings of our model to the canonical fixed effects model, we find an even larger reduction of approximately 87% for middle managers and 80% for upper managers. The fixed effects models lead to substantially larger estimates of heterogeneity in managerial quality than each of the random effects specifications, even those without match effects. A similar effect is evident in the placebo data, which illustrates that the ‘shrinkage’ characteristic found in the random effects approach is a generic feature and not driven by the nature of the dataset. Note that the fixed effects estimates would imply that some managers exert an impact on firm performance, which is three to four times larger than any accepted estimate of player influence.¹⁴ By comparison, the random effects estimates suggest that the impact of the most productive managers is comparable to the impact of star players, which we view as more realistic.

In the full model, the heterogeneity in team effects is not significantly larger in the real data than in the placebo, which suggests that team-specific factors exert less influence on performance than managerial and match effects. The polynomial of spell duration is not significant in any specification. Hence, match quality does not increase as managers cooperate longer, which opposes the finding of Huckman et al. (2009) who show that team familiarity enhances performance.¹⁵

¹⁴ For example, Brian Cashman’s FE estimate (12.68%) implies an additional 20.5 wins or an open market value of approximately \$103m per season (Cameron, 2014).

¹⁵ The difference between our results and those of Huckman et al. (2009) may be related to the setting in which team performance is examined. As compared to our setting, their teams are (a) larger (average size of 14 vs. 2) and (b) shorter lived (average 190 days vs. 661), which implies workers switch teams much more frequently. This makes it more likely that familiarity improves performance in their setting, because it may (a) increase the willingness to engage with (many) otherwise unfamiliar group members, (b) reduce the cost of coordination by establishing routines between workers that carry over to subsequent projects and (c) improve team members’ awareness of the capabilities of other members. In our setting these channels are less important, because it is clear from the onset that cooperation will be long-lasting, knowledge exchange will be intense and routines have to be established.

5.2.2. Which Individual Managers Generate the Largest Impacts?

Next, we discuss the estimates for some of the top performing managers in more detail.¹⁶ Among middle managers, Lou Piniella stands out by generating a positive run differential of 0.185 and a 2.10% increase in wins as compared to the average manager in the dataset. Only nine coaches (Buck Showalter, Bobby Cox, Larry Dierker and Jack McKeon, among others) produce run differential impacts greater than 0.10. Outside this elite group, there is relatively little difference among the vast majority of coaches in the industry. Thus, teams may find it hard to successfully replace their incumbent coach with someone who can deliver appreciable performance gains, which would explain why most coach firings are perceived to generate little positive long-term effects.

Among GMs, Brian Cashman is the most productive manager for both outcomes. We estimate that he generates approximately 5.5 wins per season more than the average GM. Since an All-Star quality player contributes approximately five wins per season relative to a replacement level player, Cashman's worth to a team is roughly on par with the value of a star player. However, a conservative estimate of the open market salary cost for an All-Star player lies around \$5 million per win or \$25 million per season (Cameron, 2014). Given that Cashman, one of the highest paid GMs, is reported to earn 'only' \$3 million per year, a team may generate a substantial surplus by hiring an elite GM. Along with Cashman, Billy Beane and Theo Epstein also generate approximately five additional wins per season. Followers of MLB will notice that Beane and Epstein were at the forefront of the sport's adoption of analytical techniques in decision making. In that sense, these results underline the power of quantitative methods to support managerial decision making.¹⁷ Moreover, both Cashman (New York Yankees) and Beane (Oakland A's) still work for their first employers and are among the highest paid GMs in the industry. It appears that their respective teams understand the edge they possess by employing these managers.

5.2.3. Robustness of Manager and Match Quality Estimates

As a first robustness check, we calculate the correlation coefficients between the individual coach and GM estimates from the various model specifications reported in Table 4.¹⁸ We find that these are positively correlated (significant at the 1% level) across all specifications. As such, the difference in managerial impacts between different models is primarily a scaling issue. All managers have a lower estimated impact

¹⁶ We provide estimates for the full top ten in Table A5 in appendix.

¹⁷ Specifically for GMs, we confirm that the 'Moneyball' managers (Lewis, 2004) outperform their peers, as argued by Hakes and Sauer (2006) and Wolfe et al. (2006). Our results diverge from Goff (2013), which is due to differences both in input controls and methodology. See Smart and Wolfe (2003) and Smart et al. (2010) for further work on the impact of GMs in MLB.

¹⁸ We provide further details on all robustness checks in the second section of the online appendix.

in the random effects approach with match controls regardless of their place in the ranking, yet each manager's position in the ranking is similar.

A potential concern with our estimates of match quality is that managers may reduce their efforts shortly before an anticipated move and then increase their efforts again after the move. Our estimates might then attribute this effort effect to the match quality of the employment spells and we would overestimate the dispersion of match qualities (Jackson, 2013). To assess this, we test whether managerial performance in the year before and after a move is different from performance in the two months just before moving, when the manager may plausibly anticipate the future move. We find no statistically significant evidence that manager performance in earlier months is better than in the two months immediately prior to a move. This indicates managers do not reduce efforts in response to future moves. However, manager performance improves after the move, which is consistent with managers moving from less to more productive matches, and again suggests that match quality is empirically relevant.

A final assumption we test is that labor market sorting is idiosyncratic to match quality. Lazear et al. (2015) suggest to test this assumption by examining the population average of the mean estimated match effects within each individual. If these means average to zero, the underlying assumption is not violated, whereas significantly positive (negative) means could indicate sorting based on unobserved match quality. Performing this test on our results produces average mean match quality estimates which are not significantly different from zero.¹⁹ Hence, we find no violation of this assumption.

5.3. Analyzing Match and Manager Qualities

5.3.1. Are High-Quality Managers Matching with High-Quality Co-Workers?

A central prediction in the literature on matching between employers and employees is that high-quality firms should attract high-quality workers. In our setting it seems natural to ask whether high-quality managers match with each other across organizational layers, or, in other words, that there is positive assortative matching. To explore this issue, we estimate the following models:

$$\hat{\mu}_{imu} = \alpha_u \hat{v}_u + \alpha_i \hat{\omega}_i + \alpha_l \tilde{l}_i + \beta_m X_m + \beta_u X_u + \varepsilon_{imu} \quad (5)$$

$$\hat{v}_{imu} = \alpha_m \hat{\mu}_m + \alpha_i \hat{\omega}_i + \alpha_l \tilde{l}_i + \beta_m X_m + \beta_u X_u + \varepsilon_{imu}. \quad (6)$$

¹⁹ For example, the values for the RE model without tenure polynomial are 0.006 (s.e.= 0.001) for GMs, 0.003 (s.e.= 0.012) for coaches, and 0.011 (s.e.= 0.013) for teams.

In equations (5) and (6), we analyze the data at the level of the individual match between coach m , GM u and team i .²⁰ The variables $\hat{\omega}_i$, $\hat{v}_u$ and $\hat{\mu}_m$ represent the standardized estimates of firm, upper and middle manager qualities, respectively. We include the log of total team payroll relative to the yearly industry average ($\tilde{l}_i$) to assess the quality of the firm workforce. Two vectors of control variables, X_m and X_u , contain observable manager characteristics (education, age, playing and management experience) for the middle and upper manager.

<Insert Table 5 around here>

In Table 5, we report OLS estimates with bootstrapped standard errors for equations (5) and (6). We report results for the random effects approach estimated without the tenure polynomial, but the results hold for each variation of the run difference and win/loss models. In the first column, we show a univariate regression of the GM on the coach effects. In further columns we gradually add more controls.

We find positive and significant results in all specifications, which clearly supports the notion of assortative matching between middle and upper managers.²¹ Further, Table 5 displays a positive association between firm and manager quality and a significant positive relationship between relative worker quality (as measured by payroll) and managerial effects. Taken together, these results indicate that better managers tend to work together at more efficient teams, who invest more in playing talent. This suggests a large degree of complementarity in ability across organizational layers, both among managers and between managers and their workers.

5.3.2. Is Match Quality Related to Manager Quality?

High-quality managers tend to work together, but do they also cooperate effectively such that they generate high match quality? To investigate this question, we estimate the following regression:

$$\hat{\sigma}_{imu} = \alpha_u \hat{v}_u + \alpha_m \hat{\mu}_m + \alpha_i \hat{\omega}_i + \alpha_l \tilde{l}_i + \beta_m X_m + \beta_u X_u + \varepsilon_{imu}. \quad (7)$$

The dependent variable in (7) is the standardized estimate of match quality ($\hat{\sigma}_{imu}$) for the pairing between GM u and coach m at team i . We are primarily interested in the estimates of α_m and α_u – the correlation between match quality and managerial quality. The control variables are the same as in equations (5) and (6), but we add an indicator for a GM who is incumbent at the firm when the coach is hired. The GM clearly

²⁰ Here we restrict our sample to spells for which we have at least 65 game observations (slightly less than half a season), which leads to a sample of 210 spells. To aid readability, we suppress the subscript i for the upper and middle manager in these and subsequent specifications, where only one team is present.

²¹ This finding does not extend to the FE estimates, as these indicate negative associations between manager effects. The explanatory power for these specifications is far below the RE models, suggesting a weaker association between estimated manager effects.

has substantial influence on the hiring of the coach, so we would expect that an entrenched GM achieves a higher match quality with a coach they have (partly) chosen, in comparison to a coach they inherited.

<Insert Table 6 around here>

Table 6 shows that all estimates we obtain for α_m and α_u are positive, highly significant and robust to the inclusion of personal characteristics. This clearly indicates that higher quality managers are associated with higher match quality. In other words, more able managers tend to cooperate better. Table 6 is again consistent with the view that managerial skills at different organizational levels are complements, rather than substitutes. The cooperation between talented managers need not result in a destructive war of egos for control over firm policies or strategy. On the contrary, talented managers excel even more when working with other talented managers. An alternative explanation may be that manager skills are positively correlated to matching ability. Following this reasoning, better managers could be simply more capable of recognizing and obtaining (e.g., through successful networking) superior cooperation opportunities. We finally find that the “GM choice” indicator is positive, but not significant, so we do not interpret the sign of this coefficient.

5.3.3. Is Match Quality Related to Differences in Managers’ Personal Backgrounds?

If manager human capital shapes the way managers interact, manager pairings with heterogeneous personal backgrounds may find it harder to achieve high match quality. We therefore examine how match quality relates to differences between the personal backgrounds of both managers. The labor economics literature argues that “mismatched” workers earn lower wages, experience higher job dissatisfaction and quicker turnover (Allen & van der Velden, 2001), which suggests they fail to make productive use of their human capital. In contrast, researchers in strategic management emphasize that a diverse management team may benefit firm performance as it could help to avoid group think (Carpenter, 2002).

To analyze this issue, we estimate the following model of team performance (y_{gijt}):

$$y_{gijt} = \gamma_{gi} - \gamma_{gj} + \beta_l(l_{git} - l_{gjt}) + \beta_k(k_{it} - k_{jt}) + \omega_i - \omega_j + \beta_m(X_{gm_i} - X_{gm_j}) + \beta_u(X_{gu_i} - X_{gu_j}) + \beta_{mu}M_{gimu} - \beta_{mu}M_{gjmu} + \varepsilon_{gijt}. \quad (8)$$

As before, we control for home advantage (γ_{gi}), labor (l_{git}), capital (k_{it}) and team fixed effects (ω_i). Instead of the manager and match effects, we insert personal characteristics of the middle and upper manager from Table 2 (X_{gm_i} and X_{gu_i}) and measure the degree of similarity between the upper and middle manager with respect to these characteristics (M_{gimu}). We include education and playing

experience as categorical variables, with their respective reference category defined as “never attended college” and “non-professional player”. We also use an indicator variable to capture whether a manager has a minority background and add the log and log square of age and managerial experience. In the M_{gimu} vector, we include separate dummy variables, which identify whether the education level between both managers differs by zero, one or two categories.²² Hence, a college educated GM working with a coach that never attended college, would differ by two categories in education. We use a similar approach for playing experience, although we allow for up to three categories of difference. As final elements of M_{gimu} we enter the log of the absolute value of the difference in age and experience between both managers. We estimate specification (8) using OLS and cluster standard errors at the level of the team-season matchup. Since we are not bound by the identification of manager and match effects, we use the full data sample for this analysis.

<Insert Table 7 around here>

In Table 7 we report our estimates for the elements of M_{gimu} ²³ (i.e., the matching variables in equation (8)). We find that a strong mismatch in education (two categories difference) is associated with lower match quality. This result supports existing work on the importance of personal backgrounds in forming social ties (e.g., Reagans, 2005). It also speaks to the idea that managers need to be able to “communicate effectively with each other,” which may be harder when educational backgrounds are very distinct. For example, think about the potential conflict between an analytically minded, college-educated GM and a non-educated coach who subscribes to a traditional approach. We also note that a small mismatch in playing experience has a negative effect of similar magnitude. Differences in age and managerial experience are not correlated with lower joint managerial performance. In contrast to Carpenter (2002), we fail to document any significant positive effect from diversity in a manager pairing.²⁴ We interpret this positive effect of homogeneity in personal background as further evidence that match quality is partly derived from manager human capital.

²² For this purpose, we drop the distinction between a top 100 vs. regular college.

²³ We refer to the online appendix for full results on other variables in the model.

²⁴ In the appendix we show full results for equation (8) and further examine the robustness of these results. Our results generally support the interpretation of manager quality as a proxy for managerial human capital, since both technical knowledge and education contribute to a manager’s ability. These findings are in line with the expert leadership hypothesis (Goodall et al., 2011; Goodall & Pogrebna, 2015) and the positive effect of education for managers in other industries (e.g., Chevalier and Ellison, 1999), but contradict the negative education effects in Mair (2005).

5.3.4. Is Match Quality Related to Differences in Strategic Approach?

To confirm that productive cooperation is driving match quality, we would ideally observe how well managers cooperate during their joint employment spells. This is not feasible in practice, however. Instead, we investigate the link between match quality and the discrepancy in the (game) strategic approach used by both managers. First, we determine the strategic approach of each manager before their joint employment spell ($\bar{s}_{tm}$ and $\bar{s}_{tu}$) by calculating their average use of a given strategy up until the time t , when the employment spell starts. We hypothesize that managers with more divergent prior approaches may find it harder to coordinate on employing specific game strategies, which could result in lower match quality. This effect may (partly) be mitigated if the middle manager is willing to adapt their strategic approach to overcome the initial disconnect. We proxy this strategic flexibility through the absolute difference between the average game strategy applied during the joint employment spell ($\bar{s}_{imu}$) and the coach's prior approach. Unfortunately, this analysis can only be conducted on employment spells in which both managers have prior managerial experience, which limits the available data points. To preserve enough statistical power, we specify the following econometric model of match quality ($\hat{o}_{imu}$):

$$\hat{o}_{imu} = \alpha_0 + \alpha_s |\bar{s}_{tu} - \bar{s}_{tm}| + \alpha_f |\bar{s}_{imu} - \bar{s}_{tm}| * |\bar{s}_{tu} - \bar{s}_{tm}| + \varepsilon_{imu}. \quad (9)$$

We use OLS to estimate the model of equation (9) for two offensive strategies – sacrificing an out to advance a baserunner (sacrifice hits) and stolen base attempts; as well as two defensive strategies – incurring errors and issuing intentional walks. We refer to Table 2 for summary statistics by manager.

<Insert Table 8 around here>

As expected, Table 8 illustrates that lower match quality is found in management pairings with a larger mismatch in their prior tendency to sacrifice an out and commit errors. The interaction of each variable with the coach's flexibility is also positive and significant.²⁵ For the other two game strategy variables, we fail to uncover significant results. In combination, these findings are consistent with the idea that cooperation among managers explains match quality. Given the limited number of employment spells in this analysis, we refrain from drawing overly strong conclusions.

5.3.5. Are Well Matched Spells Longer?

Finally, we ask whether match quality has an effect on labor market outcomes. We obviously expect well-matched managers to have an incentive to sustain their current cooperation. Likewise, a team owner

²⁵ We uncover no such effect for the GM's flexibility.

will have little to gain in breaking up a successful partnership. To test for this, we regress match duration (d_{imu}) on our estimates of match quality. We insert controls for manager and firm qualities, player payroll and match on observable manager characteristics. This results in estimating equation (10),

$$d_{imu} = \alpha_s \hat{\sigma}_{mui} + \alpha_u \hat{v}_u + \alpha_m \hat{\mu}_m + \alpha_i \hat{w}_i + \alpha_l \hat{l}_{it} + \beta_{mu} M_{imu} + \varepsilon_{imu}. \quad (10)$$

We measure the total duration of a spell as the log number of games played by team i under the manager pair.²⁶ We estimate the model in (10) using OLS with bootstrapped standard errors.²⁷

<Insert Table 9 around here>

As shown in Table 9, we find a significant and positive estimate for match quality in all specifications. Our results lend strong support to the hypothesis that higher quality spells last longer and as such, helps to validate our estimation of match quality. We further infer that coaching quality correlates more strongly to spell duration than GM quality, which never significantly impacts duration. Hence, underperforming middle managers are held accountable sooner than upper managers. This is intuitive, since the ability to terminate a spell typically lies with the GM. However, the impact of match quality on spell duration significantly outweighs the impact of the individual manager quality.

6. Discussion and Conclusion

Through our analyses, we demonstrate the importance of match quality as an economically meaningful and statistically significant driver of heterogeneity in firm performance. We argue therefore that both managerial ability and the match quality between managers are economically relevant components of firm performance. In addition, accounting for match quality significantly reduces the estimated heterogeneity in manager ability for both upper and middle managers.

In analyzing the relationship between manager characteristics, firm characteristics and match quality, we uncover that match quality does not improve along with the duration of an employment spell. This indicates that it is difficult for management pairings to improve cooperation through “learning”. Alternatively, this suggests that high performing management matches appear to be largely innate in nature. Further, we document a positive association between manager and firm quality as well as between managerial qualities across hierarchical organizational levels.

²⁶ This measures spell length more accurately than calendar days, because days during the MLB off-season (October-March) are not comparable to days during the season (April-September).

²⁷ When we repeat this analysis using a Cox hazard model, we find equivalent results.

In terms of personal backgrounds, a large mismatch in education between the upper and middle manager correlates with lower match quality. Even though high-end technical and management experience are associated with better manager performance, we do not find robust effects for matching across management levels on these characteristics. We find negative returns for pairings with conflicting prior strategic approaches at the production level, but this effect can be mitigated when a middle level manager adapts to the approach of the upper manager. Together, these findings suggest that productive cooperation across organizational levels may be an important antecedent of match quality.

Finally, we investigate the role of match and manager quality on employment spell duration. As expected, we find spells with low match quality to be shorter lived. Thus, cooperation (achieved via match quality) drives firm performance.

The implications of this study extend beyond the current context. Our results suggest that when managers work together across hierarchical levels – a characteristic of firms in most industries – hiring decisions should not be based solely on the ability level of the individual manager. Instead, firms should consider expected match quality when hiring upper and middle managers. Given that managers in high-quality matches share similar educational backgrounds and have common strategic approaches, we suggest that compatibility may largely be driven by elements such as effective communication, collaboration, and strategy alignment.

While the current paper focuses on the match between upper and middle managers, the results imply that matching between managers across other levels of the firm may also be economically meaningful. For example, future research may consider key relationships applicable to all firms, such as the match quality between the CEO and the top management team. For firms focused on research and development intensive products, the match between managers leading the research, engineering and sales units could be equally pertinent. Similarly, for manufacturing firms, the match between managers at the product development and operational levels may be more germane. As such, future research may investigate whether the findings in this paper transfer to other management pairings within modern multilevel firms across a variety of industries.

References

- Abowd, J. M., Kramarz, F., & Margolis, D. M. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2), 251-333.
- Aime, F., Johnson, S., Ridge, J. W., & Hill, A. D. (2010). The routine may be stable but the advantage is not: competitive implications of key employee mobility. *Strategic Management Journal*, 31, 75-87.
- Allen, J., & van der Velden, R. (2001). Educational Mismatches versus Skill Mismatches: Effects on Wages, Job Satisfaction and On-the-Job Search. *Oxford Economic Papers*, 53(3), 434-452.
- Artz, B. M., Goodall, A. H., & Oswald, A. J. (2017). Boss Competence and Worker Well-being. *ILR Review*, 70(2), 419-450.
- Bandiera, O., Guiso, L., Prat, A., & Sadun, R. (2015). Matching Firms, Managers and Incentives. *Journal of Labor Economics*, 33(3).
- Bertrand, M., & Schoar, A. (2003). Managing With Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics*, CXVIII(4), 1169-1208.
- Bloom, M. (1999). The Performance Effects of Pay Dispersion on Individuals and Organizations. *Academy of Management Journal*, 42(1), 25-40.
- Cameron, D. (2014, March 4). *The Cost of a Win in the 2014 Off-Season*. Retrieved from FanGraphs: <http://www.fangraphs.com/blogs/the-cost-of-a-win-in-the-2014-off-season/>
- Carpenter, M. A. (2002). The implications of strategy and social context for the relationship between top management team heterogeneity and firm performance. *Strategic Management Journal*, 23(3), 275-284.
- Chatterji, A. K., de Figueiredo, R. J., & Rawley, E. (2016). Learning on the Job? Employee Mobility in the Asset Management Industry. *Management Science*, 62(10), 2804-2819.
- Chevalier, J., & Ellison, G. (1999). Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance. *The Journal of Finance*, LIV(3), 875-899.
- Crocker, A., & Eckardt, R. (2014). A Multilevel Investigation of Individual- and Unit-Level Human Capital Complimentarities. *Journal of Management*, 40(2), 509-530.
- Depken, C. A. (2000). Wage Disparity and Team Performance: Evidence from Major League Baseball. *Economics Letters*, 67, 87-92.
- Falato, A., Li, D., & Milbourn, T. (2015). Which Skills Matter in the Market for CEOs? Evidence from Pay for CEO Credentials. *Management Science*, 61(12), 2845-2869.
- Fitza, M. A. (2014). The use of variance decomposition in the investigation of CEO effects: How large must the CEO effect be to rule out chance? *Strategic Management Journal*, 35(12), 1839-1852.
- Goff, B. (2013). Contributions of Managerial Levels: Comparing MLB and NFL. *Managerial and Decision Economics*, 34, 428-436.

- Goldfarb, A., & Xiao, M. (2011). Who thinks about competition? Managerial ability and strategic entry in US local telephone markets. *American Economic Review*, 101, 3130-3161.
- Goodall, A. H., & Pogrebna, G. (2015). Expert leaders in a fast-moving environment. *The Leadership Quarterly*, 26, 123-142. doi:10.1016/j.leaqua.2014.07.009
- Goodall, A. H., Kahn, L. M., & Oswald, A. J. (2011). Why do leaders matter? A study of expert knowledge in a superstar setting. *Journal of Economic Behavior & Organization*, 77, 265-284.
- Guadalupe, M., Hongyi, L., & Wulf, J. (2014). Who lives in the C-suite? Organizational Structure and the Division of Labor in Top Management. *Management Science*, 60(4), 824-844.
- Hakes, J. K., & Sauer, R. D. (2006). An Economic Evaluation of the Moneyball Hypothesis. *Journal of Economic Perspectives*, 20(6), 173-185.
- Hambrick, D. C. (2007). Upper Echelons Theory: An Update. *Academy of Management Review*, 32(2), 334-343.
- Hambrick, D. C., & Mason, P. A. (1984). Upper Echelons: The Organization as a Reflection of Its Top Managers. *The Academy of Management Review*, 9(2), 193-206.
- Hendricks, K. B., Hora, M., & Singhal, V. R. (2014). An Empirical Investigation on the Appointments of Supply Chain and Operations Management Executives. *Management Science*, 61(7), 1562 - 1583.
- Herbst, D., & Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, 350(6260), 545-549.
- Hitt, M. A., Beamish, P. W., Jackson, S. E., & Mathieu, J. E. (2007). Building Theoretical and Empirical Bridges Across Levels: Multilevel Research in Management. *Academy of Management Journal*, 1385-1399.
- Huckman, R. S., Staats, B. R., & Upton, D. M. (2009). Team Familiarity, Role Experience, and Performance: Evidence from Indian Software Services. *Management Science*, 55(1), 85-100.
- Jackson, K. C. (2013). Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *The Review of Economics and Statistics*, 95(4), 1096-1116.
- Juravich, M., Salaga, S., & Babiak, K. (2017). Upper Echelons in Professional Sport: The Impact of NBA General Managers on Team Performance. *Journal of Sport Management*, 466-479.
- Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *Journal of Economic Perspectives*, 14(3), 75-94.
- Kaplan, S. N., Klebanov, M. M., & Sorensen, M. (2012). Which CEO Characteristics and Abilities Matter? *The Journal of Finance*, LXVII(3), 973-1007.
- Lazear, E. P., Shaw, K. L., & Stanton, C. T. (2015). The Value of Bosses. *Journal of Labor Economics*, 33(4), 823-861.
- Lewis, M. (2004). *Moneyball: the art of winning an unfair game*. W. W. Norton & Company.

- Mair, J. (2005). Exploring the Determinants of Unit Performance: The Role of Middle Managers in Stimulating Profit Growth. *Group & Organization Management*, 30, 263-288.
- Malmendier, U., Tate, G., & Yan, J. (2011). Overconfidence and Early-Life Experiences: The Effect of Managerial Traits on Corporate Financial Policies. *The Journal of Finance*, LXVI(5), 1687-1732.
- Owan, H., Takahashi, S., Tsuru, T., & Uehara, K. (2014). *Finding good managers: an econometric case study of a large Japanese auto dealership*. Working paper.
- Pan, Y. (2017). The Determinants and Impact of Executive-Firm Matches. *Management Science*, 63(1), 185-200.
- Peeters, T., & Szymanski, S. (2014). Financial Fair Play in European football. *Economic Policy*, 29(78), 343-390.
- Ployhart, R. E., & Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *The Academy of Management Review*, 36(1), 127-150.
- Ployhart, R. E., Nyberg, A. J., Reilly, G., & Maltarich, M. A. (2014). Human Capital Is Dead; Long Live Human Capital Resources! *Journal of Management*, 40(2), 371-398.
- Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876–2003). *Journal of Sport Sciences*, 23(5), 337-350.
- Raes, A. M., Heijltjes, M. G., Glunk, U., & Roe, R. A. (2011). The Interface of the Top Management Team and Middle Managers: A Process Model. *Academy of Management Review*, 36(1), 102-126.
- Reagans, R. (2005). Preferences, identity, and competition: Predicting tie strength from demographic data. *Management Science*, 51(9), 1374-1383.
- Siebert, S. W., & Zubanov, N. (2010). Management Economics in a Large Retail Company. *Management Science*, 56(8), 1398-1414.
- Smart, D. L., & Wolfe, R. A. (2003). The Contribution of Leadership and Human Resources to Organizational Success: An Empirical Assessment of Performance in Major League Baseball. *European Sport Management Quarterly*, 3, 165-188.
- Smart, D. L., Winfree, J., & Wolfe, R. (2010). Major League Baseball Managers: Do They Matter? *Journal of Sport Management*, 22(3), 303-321.
- Staw, B. M., & Hoang, H. (1995). Sunk Costs in the NBA: Why Draft Order Affects Playing Time and Survival in Professional Basketball. *Administrative Science Quarterly*, 40, 474-494.
- Szymanski, S. (2003). The Economic Design of Sporting Contests. *Journal of Economic Literature*, 41(4), 1137-1187.
- Terviö, M. (2008). The Difference That CEOs Make: An Assignment Model Approach. *American Economic Review*, 98(3), 642-668.
- Wolfe, R., Wright, P. M., & Smart, D. L. (2006). Radical HRM Innovation and Competitive Advantage: the Moneyball Story. *Human Resource Management*, 45(1), 111-145.

- Wooldridge, B., & Floyd, S. W. (1990). The strategy process, middle management involvement, and organizational performance. *Strategic Management Journal*, 11(3), 231-241.
- Wooldridge, B., Schmid, T., & Floyd, S. W. (2008). The Middle Management Perspective on Strategy Process: Contributions, Synthesis, and Future Research. *Journal of Management*, 34(6), 1190-1221.
- Wright, P. M., & McMahan, G. C. (2011). Exploring human capital: putting 'human' back into strategic human resource management. *Human Resource Management Journal*, 21(2), 93-104.

Tables and Figures

Table 1: Summary Statistics – Estimation Sample

	Obs.	Mean	S.D.	Min	Max
Output					
Run difference	84604	0	4.365	-27	27
Game result	84604	0.5	0.5	0	1
Inputs					
Team payroll ('12 m\$)	84604	76.0	37.6	13.0	244.0
S.D. team payroll ('12 m\$)	84604	3.27	1.56	0.31	9.43
Start pitcher salary ('12 m\$)	84604	2.82	3.77	0.07	24.3
MSA home value ('12 m\$)	84604	0.36	0.20	0.10	1.50
Stadium age	84604	27.8	25.5	1	101
Fixed Effects					
Team id	84604	14.4	8	1	28
GM id	84604	38.6	22.5	1	78
Coach id	84604	50.2	30.4	1	106
Spell id	84604	152.2	83.9	1	291

Notes: This table presents summary statistics for the estimation sample at the individual game level. The sources for data are www.shrpsports.com/mlb (game results), content.usatoday.com/sportsdata/baseball/mlb/salaries (payroll data), www.retrosheet.org (line-ups), www.ballparks.com (stadium data), www.lincolnst.edu/research-data/data (home values), www.census.gov/programs-surveys/metro-micro.html (MSA level data) and www.baseball-reference.com/ (manager data).

Table 2: Summary Statistics – Manager Personal Characteristics, Career Statistics & Strategic Approach

	Coaches					General Managers				
	Obs.	Mean	S.D.	Min	Max	Obs.	Mean	S.D.	Min	Max
Personal Characteristics										
College attended	105	0.733	0.444	0	1	74	0.865	0.344	0	1
College graduated	105	0.429	0.497	0	1	74	0.797	0.405	0	1
College top 100						74	0.189	0.394	0	1
Professional player	105	0.971	0.167	0	1	76	0.421	0.497	0	1
MLB player	105	0.781	0.416	0	1	76	0.197	0.401	0	1
MLB All-Star	105	0.324	0.470	0	1	76	0.053	0.225	0	1
Minority	105	0.152	0.361	0	1	76	0.079	0.271	0	1
Age (days)	105	18477	2196	13017	24991	75	17426	2920	11655	24767
Career Characteristics										
Avg. run difference	105	-0.123	0.460	-1.690	0.738	77	-0.029	0.387	-0.910	0.909
Avg. win%	105	48.6%	4.72%	35.5%	57.3%	77	49.7%	4.00%	39.9%	60.3%
Experience (days)	105	2080	2253	88	9517	76	3625	2068	728	9865
Strategic Approach										
Avg. sacrificed hits	105	0.020	0.194	-0.422	0.439	77	0.047	0.129	-0.326	0.370
Avg. steal attempts	105	0.160	0.379	-0.483	1.659	77	0.165	0.234	-0.230	0.942
Avg. errors	105	0.114	0.234	-0.248	0.794	77	0.110	0.154	-0.090	0.697
Avg. intentional walk	105	0.021	0.149	-0.259	0.378	77	0.041	0.086	-0.136	0.251

Notes: This table shows summary statistics of managers' personal characteristics, career characteristics and strategic approach at the level of the individual manager in the estimation sample. The main sources for data are www.baseball-reference.com/ and www.retrosheet.org/. We consulted various other websites (e.g., official club sites) if data were missing.

Table 3: 1st Stage Regression Results – Determinants of Game Level Run Difference

Dep. Var.: Run Difference	No FE	Team FE	Coach FE	GM FE	All FE	Spell FE
Home advantage	0.217*** (0.042)	0.217*** (0.042)	0.217*** (0.042)	0.217*** (0.042)	0.217*** (0.042)	0.214*** (0.042)
Team payroll	0.595*** (0.086)	0.495*** (0.095)	0.291*** (0.111)	0.334*** (0.108)	0.271*** (0.124)	0.270*** (0.132)
S.D. team payroll	-0.229*** (0.087)	-0.274*** (0.091)	-0.261*** (0.096)	-0.211** (0.100)	-0.157 (0.107)	-0.098 (0.112)
Start pitcher salary	0.141*** (0.012)	0.134*** (0.012)	0.129*** (0.012)	0.128*** (0.012)	0.123*** (0.012)	0.121*** (0.012)
MSA home value	0.117*** (0.037)	0.215 (0.131)	0.421** (0.187)	0.038 (0.185)	0.188 (0.244)	0.411 (0.283)
Stadium age	-0.022 (0.015)	-0.028 (0.019)	-0.044* (0.023)	-0.014 (0.022)	-0.007 (0.027)	0.008 (0.029)
Constant	-0.108*** (0.021)	-0.108*** (0.021)	-0.108*** (0.021)	-0.108*** (0.021)	-0.109*** (0.021)	-0.107*** (0.021)
Team FE	No	Yes	Yes	Yes	Yes	No
F-test	-	5.117***	2.457***	2.195***	1.643**	-
Coach FE	No	No	Yes	No	Yes	No
F-test	-	-	3.699***	-	2.561***	-
GM FE	No	No	No	Yes	Yes	No
F-test	-	-	-	4.467***	2.836***	-
Spell FE	No	No	No	No	No	Yes
Observations	84,604	84,604	84,604	84,604	84,604	84,604
R-squared	0.010	0.014	0.022	0.022	0.027	0.030
Adj. R-squared	0.010	0.013	0.019	0.019	0.022	0.023

Notes: This table shows results for a linear regression model explaining the run difference in a game as a function of the inputs used by both teams and various sets of team, manager and spell fixed effects. Standard errors are clustered at the seasonal matchup level and are shown in parentheses. *** denotes significance at 1%-level, ** significance at 5%-level, and * significance at 10%-level.

Table 4: Standard Deviation of Match, Manager and Team Effects in Real and Placebo Data Models

Level	Dependent Variable	Std. Dev. Random Effects			Std. Dev. Fixed Effects		# Effects
Match	Real run dif.		0.165	0.165			210
	<i>Placebo run dif.</i>		<i>0.077</i>	<i>0.079</i>			
	Real win%		1.15%	1.17%			
	<i>Placebo win%</i>		<i>0.000</i>	<i>0.000</i>			
Coach	Real run dif.	0.160	0.065	0.058	0.526	0.537	105
	<i>Placebo run dif.</i>	<i>0.056</i>	<i>0.009</i>	<i>0.007</i>	<i>0.223</i>	<i>0.225</i>	
	Real win%	1.62%	0.95%	0.93%	5.49%	5.68%	
	<i>Placebo win%</i>	<i>0.00%</i>	<i>0.00%</i>	<i>0.00%</i>	<i>4.25%</i>	<i>4.28%</i>	
GM	Real run dif.	0.197	0.146	0.146	0.736	0.741	77
	<i>Placebo run dif.</i>	<i>0.045</i>	<i>0.009</i>	<i>0.007</i>	<i>0.313</i>	<i>0.320</i>	
	Real win%	1.74%	1.28%	1.29%	5.84%	5.99%	
	<i>Placebo win%</i>	<i>0.00%</i>	<i>0.00%</i>	<i>0.00%</i>	<i>6.15%</i>	<i>6.21%</i>	
Team	Real run dif.	0.108	0.020	0.022	0.681	0.689	28
	<i>Placebo run dif.</i>	<i>0.062</i>	<i>0.063</i>	<i>0.017</i>	<i>0.338</i>	<i>0.345</i>	
	Real win%	0.68%	0.28%	0.29%	5.08%	5.53%	
	<i>Placebo win%</i>	<i>0.00%</i>	<i>0.00%</i>	<i>0.00%</i>	<i>6.67%</i>	<i>6.71%</i>	
Controls for:							
Match quality		No	Yes	Yes	No	No	
Spell duration polynomial		No	No	Yes	No	Yes	

Notes: This table shows the standard deviation of the estimated match, manager and team effects for different versions of the team production model. We show results for the random and fixed effects models with and without the spell duration polynomial and match effects. The numbers in italics refer to the average standard deviation found in 500 placebo datasets, where the dependent variable is randomly reassigned across observations.

Table 5: Regression Results – Are High-Quality Managers Matching with High-Quality Co-Workers?

Dependent Variable:	Run Difference					Win %				
	GM Effect			Coach Effect		GM Effect			Coach Effect	
Coach effect	0.307*** (0.079)	0.197*** (0.073)	0.221*** (0.074)			0.266*** (0.063)	0.181*** (0.063)	0.191*** (0.077)		
GM effect				0.219*** (0.070)	0.221*** (0.080)				0.205*** (0.060)	0.208*** (0.068)
Team effect		0.296*** (0.052)	0.197*** (0.068)	0.107 (0.071)	0.178** (0.077)		0.289*** (0.074)	0.176** (0.075)	0.084 (0.066)	0.138* (0.085)
Rel. payroll		0.683*** (0.255)	0.745*** (0.221)	0.588*** (0.221)	0.451** (0.220)		0.751*** (0.202)	0.800*** (0.215)	0.422* (0.234)	0.353 (0.230)
GM/coach pers. char.	no	no	yes	no	yes	no	no	yes	no	yes
Observations	210	210	204	210	204	210	210	204	210	204
R-squared	0.094	0.220	0.327	0.134	0.334	0.071	0.197	0.329	0.091	0.271
Adj. R-squared	0.090	0.209	0.262	0.122	0.269	0.066	0.186	0.263	0.078	0.200

Notes: This table presents results for a linear regression model explaining the GM (coach) effect as a function of the GM's (coach's) co-worker's effect, the team effect, the relative player payroll of the team and the GM's (coach's) personal characteristics. Personal characteristics include GM (coach) age, experience, education and professional playing experience. Bootstrapped standard errors are given in parentheses. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level.

Table 6: Regression Results – Is Match Quality Related to Manager Quality?

Dependent Variable:	Match Effect: Run Difference			Match Effect: Win%		
Coach effect	0.430*** (0.075)	0.426*** (0.078)	0.441*** (0.065)	0.414*** (0.070)	0.413*** (0.057)	0.419*** (0.081)
GM effect	0.258*** (0.056)	0.262*** (0.077)	0.280*** (0.064)	0.306*** (0.073)	0.309*** (0.060)	0.355*** (0.094)
Team effect	0.191*** (0.047)	0.188*** (0.062)	0.244*** (0.059)	0.152*** (0.054)	0.149** (0.060)	0.197*** (0.047)
Rel. player payroll	-0.410** (0.180)	-0.396** (0.198)	-0.404** (0.193)	-0.460** (0.220)	-0.453*** (0.164)	-0.474* (0.246)
GM incumbent		0.088 (0.098)	0.075 (0.139)		0.064 (0.104)	0.023 (0.113)
Constant	0.406** (0.185)	0.369* (0.208)	1.550 (4.659)	0.456** (0.226)	0.432** (0.172)	-1.518 (6.320)
GM and coach pers. char.	No	No	Yes	No	No	Yes
Observations	210	210	204	210	210	204
R-squared	0.389	0.390	0.436	0.379	0.380	0.425
Adj. R-squared	0.377	0.375	0.374	0.367	0.364	0.362

Notes: This table shows regression results for a linear model explaining the estimated match effect in an employment spell as a function of the estimated individual effects and personal characteristics of both managers in the spell. We further control for the team effect, relative player payroll and a dummy indicating the GM was incumbent at the start of the spell. The personal characteristics include GM and coach age, experience, education and professional playing experience. Bootstrapped standard errors are given in parentheses. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level.

Table 7: Regression Results – Is Performance Related to Differences in Manager Personal Background?

Dependent Variable:		Run Difference	Win%
Difference in:			
Education	1 cat.	0.079 (0.098)	0.006 (0.011)
	2 cat.	-0.155** (0.062)	-0.016** (0.007)
	3 cat.	-0.162*** (0.061)	-0.013* (0.007)
Play level	1 cat.	-0.105 (0.088)	-0.005 (0.010)
	2 cat.	-0.128 (0.133)	-0.013 (0.016)
	3 cat.	0.005 (0.017)	0.000 (0.002)
Experience		-0.001 (0.019)	-0.001 (0.002)
Age			
Controls for:			
Inputs		Yes	Yes
Levels personal char.		Yes	Yes
Team FE		Yes	Yes
Observations		104,204	104,204
R-squared		0.018	0.021
Adj. R-squared		0.017	0.020

Notes: This table shows results for a linear regression model explaining game outcomes (run difference and win/loss) as a function of the inputs used by both teams, the level and difference in personal characteristics of managers employed by both teams and team fixed effects. Categories in education refer to no college/college attended/college graduated; categories in playing experience are not a pro player/pro player/MLB player/All-Star. The difference is expressed as the number of ordered categories in which matched managers differ. The results for the level characteristics and inputs is suppressed to aid readability and can be found in appendix Table A7. Standard errors in parentheses are clustered at the seasonal matchup level. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level.

Table 8: Regression Results – Is Match Quality Related to Differences in Strategic Approach?

Dependent Variable:	Match Effect Run Difference				Match Effect Win%			
Difference sacrificed hits	-2.515*				-2.348*			
	(1.382)				(1.356)			
Flexibility * difference sacrificed hits	6.350**				6.204*			
	(3.190)				(3.129)			
Difference steal attempts		-0.719				-0.991		
		(0.817)				(0.796)		
Flexibility * difference steal attempts		0.578				0.689		
		(1.612)				(1.571)		
Difference errors			-2.614				-4.688**	
			(2.216)				(2.141)	
Flexibility * difference errors			13.35**				17.91***	
			(6.350)				(6.137)	
Difference intentional walk				0.367				0.483
				(1.303)				(1.279)
Flexibility * difference intentional walk				-2.514				-2.380
				(3.935)				(3.860)
Constant	0.192	0.120	-0.003	-0.000	0.163	0.173	0.117	-0.024
	(0.198)	(0.167)	(0.169)	(0.169)	(0.194)	(0.163)	(0.164)	(0.166)
Observations	97	97	97	97	97	97	97	97
R-squared	0.042	0.011	0.060	0.005	0.041	0.023	0.087	0.004
Adj. R-squared	0.021	0.000	0.040	0.000	0.020	0.002	0.067	0.000

Notes: This table presents results of regressing the estimated match effect in a manager pairing on the difference in the prior strategic approaches of both managers and the adaptation of the coach during the spell to the GM's prior approach interacted with the prior difference. Adjusted R-squared values below zero are reported as 0.000. Standard errors are given in parentheses. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level.

Table 9: Regression Results – Are Well Matched Spells Longer?

Spell Duration	Run Difference			Win%		
Match effect	0.219*** (0.056)	0.135* (0.077)	0.158* (0.084)	0.246*** (0.050)	0.174** (0.073)	0.178** (0.080)
Team effect		0.038 (0.056)	0.018 (0.059)		0.029 (0.069)	0.009 (0.062)
Coach effect		0.111 (0.080)	0.087 (0.065)		0.115* (0.059)	0.099 (0.063)
GM effect		0.034 (0.054)	-0.009 (0.064)		0.015 (0.050)	0.009 (0.053)
Player payroll		-0.029 (0.174)	-0.002 (0.230)		0.022 (0.221)	0.041 (0.203)
Constant	5.864*** (0.055)	5.892*** (0.174)	6.387*** (0.638)	5.868*** (0.047)	5.846*** (0.176)	6.375*** (0.652)
Difference pers. char.	No	No	Yes	No	No	Yes
Observations	186	186	180	186	186	180
R-squared	0.087	0.111	0.149	0.107	0.131	0.171
Adj. R-squared	0.082	0.086	0.088	0.102	0.107	0.112

Notes: This table shows regression results for a linear model explaining the duration of a joint employment spell expressed in games as a function of the estimated match, manager and team effects. We further control for the difference in personal characteristics between the GM and coach in the match with regard to age, experience, education and professional playing experience. We suppress these coefficients to improve readability. Bootstrapped standard errors are given in parentheses. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level.

Appendix (Online Publication Only)

A1. Robustness Checks: 1st Stage Regressions

Here, we provide additional robustness checks for the first stage results. Table A1 displays the full results for the input parameters when using the game result as dependent variable. These estimates confirm the effects found in Table 3.

Table A1: 1st Stage Regression Results – Determinants of Game Result

Game Result/Win%	No FE	Team FE	Coach FE	GM FE	All FE	Spell FE
Home advantage	0.073*** (0.005)	0.073*** (0.005)	0.073*** (0.005)	0.073*** (0.005)	0.073*** (0.005)	0.073*** (0.005)
Team payroll	0.066*** (0.010)	0.060*** (0.011)	0.041*** (0.012)	0.040*** (0.012)	0.037*** (0.014)	0.042*** (0.015)
S.D. team payroll	-0.029*** (0.010)	-0.034** (0.011)	-0.034*** (0.011)	-0.021** (0.012)	-0.024* (0.012)	-0.017 (0.013)
Pitcher salary	0.015*** (0.001)	0.014*** (0.001)	0.013*** (0.001)	0.014*** (0.001)	0.013*** (0.001)	0.013*** (0.001)
MSA home value	0.015*** (0.004)	0.028* (0.015)	0.058*** (0.021)	0.004 (0.021)	0.015 (0.028)	0.028 (0.032)
Stadium age	-0.002 (0.002)	-0.002 (0.002)	-0.005* (0.003)	0.001 (0.002)	-0.000 (0.003)	-0.001 (0.003)
Constant	0.463*** (0.002)	0.463*** (0.002)	0.463*** (0.002)	0.463*** (0.002)	0.463*** (0.002)	0.464*** (0.002)
Team FE	No	Yes	Yes	Yes	Yes	No
F-test	-	5.084***	2.928***	1.639**	0.882	-
Coach FE	No	No	Yes	No	Yes	No
F-test	-	-	3.317***	-	2.260***	-
GM FE	No	No	No	Yes	Yes	No
F-test	-	-	-	3.710***	2.175***	-
Spell FE	No	No	No	No	No	Yes
Observations	84,604	84,604	84,604	84,604	84,604	84,604
R-squared	0.014	0.017	0.025	0.024	0.028	0.031
Adj. R-squared	0.014	0.017	0.021	0.021	0.023	0.024

Notes: This table shows results for a linear regression model explaining the win-loss result in a game as a function of the inputs used by both teams and various sets of team, manager and spell fixed effects. Standard errors are clustered at the seasonal matchup level and are shown in parentheses. *** denotes significance at 1%-, ** significance at 5%- and * significance at 10%-level.

A potential problem in estimating production functions is that firms may choose adjustable inputs, such as labor, conditional on unobservable factors, such as productivity, leading to omitted variable bias. These are therefore treated as potentially endogenous. A further concern in our case is that game results have a feedback effect on payroll, for example through performance-related bonuses. To assess the robustness of our results with respect to these potential biases we estimate equation (3) with and without

instrumenting the total wage and standard deviation of wages. As instruments we employ lagged payroll values, stadium capacity and a set of market characteristics for the metropolitan statistical area (MSA) where the team is located, a time trend, and a dummy for teams playing in the American League. We then re-estimate the baseline first-stage model as in Table 3, for both dependent variables with and without instrumentation. We also test the following variations in model specification:

- a) Add a third degree polynomial of spell duration to allow match quality to evolve over time.
- b) Include coach and GM age, experience and their square terms.
- c) Interact input measures with a 'post-season', 'home game' and 'local derby' dummy.
- d) Drop the standard deviation of team payroll and starting pitcher salary.
- e) Include controls for MSA population, number of teams in the MSA, the club's tenure in the MSA and stadium capacity directly in the 1st stage.

For each specification, we calculate the corresponding second stage dependent variable, i.e., the combined spell fixed effect plus the residual for each observation. We then determine the correlation coefficient between each of these alternatives and the baseline model from Table 3. All of the resulting correlation coefficients (see Table A2) are at or above 0.98 and significant at the 1% level, which ensures that our results are not overly dependent on the exact choice of specification for the baseline model.

Table A2: Correlation Coefficients of Estimated Spell Fixed Effects for Alternative 1st Stage Specifications

Dep. Variable	IV Equivalent	Add Duration Poly.	Add GM/Coach Exp.	Interact Post-Season	Interact Home Adv.	Interact Derby Games	Drop S.D. Payroll/Pitcher Salary	Add MSA-Level Controls
Run dif.	0.999	1.000	0.996	1.000	1.000	1.000	0.999	0.987
+ IV wage	n.a.	1.000	0.998	1.000	1.000	1.000	0.998	0.986
Win%	0.999	0.999	0.998	1.000	1.000	1.000	0.999	0.983
+ IV wage	n.a.	1.000	0.999	1.000	1.000	1.000	0.999	0.980
# of employment spells		291	275	291	291	291	291	291
# of observations		84604	82102	84604	84604	84604	84604	84604

Notes: This table depicts the correlation coefficients among the spell fixed effects (= 2nd stage dependent variable) for various specifications of the 1st stage regression model. Values of 1.000 are larger than 0.999, but do not equal one. All values are significant at the 1% level.

A2. Extra Results and Robustness Checks for Match and Manager Effects

In this section, we provide more detail on the results and robustness checks described in section 5.2.

First, Table A3 reports the correlation coefficients for coach, GM and team effects across various empirical approaches.

Table A3: Correlation Coefficients of Manager and Team Effects across Estimation Methods

Correlation Individual Effects		Run Difference	Win%	Obs.
FE no tenure – FE tenure	Coach	0.9712***	0.9647***	105
	GM	0.9836***	0.9436***	77
	Team	0.9599***	0.9056***	28
FE no tenure – RE no spell, no tenure	Coach	0.7323***	0.7273***	105
	GM	0.5936***	0.6274***	77
	Team	0.2119	0.2983	28
RE no spell, no tenure – RE spell, no tenure	Coach	0.9322***	0.9702***	105
	GM	0.9515***	0.9777***	77
	Team	0.9199***	0.9585***	28
RE spell, no tenure – RE spell, tenure	Coach	0.9993***	0.9996***	105
	GM	0.9994***	0.9995***	77
	Team	0.9993***	0.9996***	28

Notes: This table shows the correlation coefficients among coach, GM and team effects estimated using varying methodologies and model specifications for the 1st stage model. *** denotes significance at 1%-level.

To perform the test described in Jackson (2013) in our setting, we isolate observations one year (160 games) before and after a manager moves, where we define a move as any change in the team-GM-coach partnership.²⁸ Next, we split the period before and after the move (320 games in total) into eight periods of 40 games. We then estimate a model which explains performance as a function of each team's inputs (as in the baseline model), moving manager (coach or GM) FEs, team FEs and a series of indicators (denoted τ_{m+k}) capturing the eight periods around the period of the move (w), with the last 40 games before the move being the reference category. The regression equation takes the following form,

$$y_{gijt} = \gamma_{gi} - \gamma_{gj} + \beta_l(l_{git} - l_{gjt}) + \beta_k(k_{it} - k_{jt}) + \omega_i + \mu_i + v_i + \sum_{k=-4}^{+4} \tau_{w+k} + \varepsilon_{gijt}. \quad (A1)$$

²⁸ We only use employment spells which attain at least 160 games in this analysis to avoid that our results would be driven by the effect of employment spells ending in sample. We experimented with alternative choices for sample selection, i.e., up to two years on each side of the move. This drastically reduces the data available for analysis, such that we present the single year results here.

We show the results of this exercise in Table A4, where estimates of the input coefficients are omitted to aid readability. Most importantly, managers do not perform significantly different in the period right before a move versus those earlier in the employment spell. The period dummies reveal few significant coefficients. We confirm this by performing an F-test on the joint significance of the pre-move periods, which shows statistical significance at the 10% level in only one of the eight specifications. The post-period reveals a somewhat different picture. Here, we find additional significant coefficients, while the F-test results suggest that these coefficients are significant in a number of specifications. This finding supports the idea that managers abandon relatively unproductive matches to form, on average, better matched pairs. We again note there is little evidence of systematic variation in match quality over the employment spell, which supports the notion that managers do not improve their cooperation as a match progresses.

Table A4: Regression Results – Does Manager Performance Change Close to Turnover?

Outcome Measure Manager Moving	Run Difference				Win %			
	Coach		GM		Coach		GM	
160-120 games pre move	0.166 (0.104)	0.208** (0.104)	0.126 (0.102)	0.174* (0.102)	0.021* (0.012)	0.025** (0.012)	0.022* (0.012)	0.028** (0.012)
120-80 games pre move	0.057 (0.104)	0.106 (0.104)	0.085 (0.103)	0.130 (0.103)	0.011 (0.012)	0.016 (0.012)	0.010 (0.012)	0.017 (0.012)
80-40 games pre move	0.000 (0.113)	0.049 (0.112)	0.168 (0.103)	0.208** (0.104)	0.004 (0.013)	0.007 (0.013)	0.009 (0.012)	0.015 (0.012)
last 40 games pre move	Reference category							
first 40 games post move	0.114 (0.113)	0.214* (0.114)	0.204* (0.108)	0.289*** (0.109)	0.027** (0.013)	0.037*** (0.013)	0.019 (0.012)	0.031** (0.012)
40-80 games post move	0.053 (0.112)	0.135 (0.116)	0.037 (0.107)	0.114 (0.109)	0.017 (0.012)	0.025* (0.013)	0.001 (0.012)	0.012 (0.012)
80-120 games post move	0.234** (0.111)	0.326*** (0.113)	0.176 (0.108)	0.248** (0.110)	0.034*** (0.012)	0.044*** (0.013)	0.023* (0.012)	0.032*** (0.012)
120-160 games post move	0.054 (0.113)	0.134 (0.114)	0.066 (0.115)	0.152 (0.117)	0.010 (0.013)	0.019 (0.014)	-0.005 (0.013)	0.007 (0.013)
Controls inputs/opp.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Manager FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Team FE	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.022	0.027	0.020	0.025	0.022	0.027	0.023	0.027
Observations	18,685	18,685	19,226	19,226	18,685	18,685	19,226	19,226
F-test prob. pre=0	0.392	0.230	0.353	0.148	0.310	0.196	0.302	0.097
F-test prob. post=0	0.289	0.068	0.269	0.067	0.057	0.009	0.116	0.028

Notes: F-test refers to joint significance of all pre-move or post-move dummy variables. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level.

Finally, Table A5 provides an overview of the top GMs and coaches in terms of estimated individual effects. We highlight some of these results in section 5.2.2 of the main body of the paper.

Table A5: Individual Manager Contributions to Run Differential and Winning Percentage

Coach Name	Run Dif.		Win%		# Manager of Year Awards
	RE	FE	RE	FE	
Lou Piniella	0.185	0.344	2.10%	6.57%	3
Buck Showalter	0.132	0.958	1.53%	8.94%	2
Bobby Cox	0.132	0.232	1.55%	-2.48%	3
Larry Dierker	0.127	0.750	1.32%	2.07%	1
Jack McKeon	0.126	0.522	2.12%	6.79%	2
Joe Torre	0.120	0.314	1.83%	1.84%	2
Fredi Gonzalez	0.116	0.352	2.15%	3.55%	0
Charlie Manuel	0.111	0.087	1.30%	1.32%	0
Tony La Russa	0.110	0.671	1.56%	8.13%	3
Davey Johnson	0.097	0.389	1.18%	5.10%	1
GM Name	Run Dif.		Win%		# Executive of Year Awards
	RE	FE	RE	FE	
Brian Cashman	0.343	1.365	3.37%	12.68%	0
Pat Gillick	0.294	0.507	2.33%	3.59%	1
Jon Daniels	0.291	0.814	1.98%	-5.32%	0
Theo Epstein	0.290	1.302	2.98%	11.86%	0
Billy Beane	0.265	1.915	3.12%	20.99%	2
John Mozeliak	0.249	-0.798	1.37%	0.23%	0
John Schuerholz	0.243	0.802	2.28%	9.65%	0
J.P. Ricciardi	0.224	1.228	1.14%	5.70%	0
Gerry Hunsicker	0.208	-0.149	1.31%	-0.76%	1
Frank Wren	0.207	0.563	1.09%	4.28%	0

Notes: This table displays the individual estimates for the ten managers with the highest random effects for run difference at each level. Random effects estimates are drawn from the fully specified model with tenure polynomial and match effects. The fixed effects estimates derive from the model with tenure polynomial. All estimates have been rescaled to make the average of all manager effects equal to zero.

A3. Additional Results: Personal Characteristics

In this section, we provide additional results for the analysis described in section 5.2.3. First, Table A6 shows the estimation results for equation (8), which are not reported in Table 7 as we highlight the most relevant of these in the main text.

We then examine the robustness of the matching results by estimating an adapted version of the match quality model in equation (7), adding both the level of, and matching on, observable characteristics (X_{mi} , X_{ui} and M_{mui} respectively), i.e.,

$$\hat{\sigma}_{imu} = \alpha_u \hat{v}_u + \alpha_m \hat{\mu}_m + \alpha_i \hat{\omega}_i + \alpha_l \widetilde{l_{it}} + \beta_m X_{mi} + \beta_u X_{ui} + \beta_{mu} M_{imu} + \varepsilon_{imu}. \quad (A2)$$

To avoid multi-collinearity issues, we enter characteristics one by one and include either our estimates of manager and team qualities ($\hat{v}_u$, $\hat{\mu}_m$ and $\hat{\omega}_i$) or the level of the observable characteristics (X_{mi} and X_{ui}). The results of this exercise, shown in Table A7, largely confirm the effect of a strong mismatch on education. We do not find support for a mismatch on playing (technical) experience, or on other observable characteristics. If anything, a mismatch in experience may yield marginally positive effects on match quality.

Table A6: Full Regression Results – Match Quality and Managers' Personal Backgrounds

	Run Difference				Win%			
	Only Level		Level + Dif. Match		Only Level		Level + Dif. Match	
Pers. Char.:	Coach	GM	Coach	GM	Coach	GM		
Professional player	0.256*	0.142***	0.386**	0.161**	0.030*	0.018***	0.041**	0.020**
	(0.153)	(0.047)	(0.161)	(0.067)	(0.017)	(0.005)	(0.018)	(0.008)
MLB player	0.126	-0.021	0.234	-0.025	0.013	-0.000	0.021	-0.000
	(0.153)	(0.063)	(0.167)	(0.097)	(0.017)	(0.007)	(0.019)	(0.011)
MLB All-Star	0.328**	0.196**	0.440**	0.120	0.034**	0.020**	0.045**	0.012
	(0.153)	(0.087)	(0.186)	(0.127)	(0.017)	(0.010)	(0.021)	(0.015)
College not grad.	0.135***	0.315***	-0.042	0.208*	0.013**	0.017	-0.003	0.009
	(0.050)	(0.096)	(0.103)	(0.114)	(0.006)	(0.011)	(0.012)	(0.013)
College graduated	0.133***	0.187***	0.051	0.159***	0.013**	0.018***	0.004	0.016**
	(0.051)	(0.057)	(0.067)	(0.058)	(0.006)	(0.007)	(0.008)	(0.007)
College top 100		0.175**		0.153**		0.022***		0.021***
		(0.068)		(0.069)		(0.008)		(0.008)
Minority	-0.069	0.140*	-0.069	0.074	-0.001	0.012	0.000	0.004
	(0.054)	(0.082)	(0.056)	(0.086)	(0.006)	(0.009)	(0.006)	(0.010)
Age	-12.141	-22.667***	-18.349	-22.959**	-0.765	-2.123**	-1.403	-2.114**
	(12.367)	(8.345)	(13.093)	(8.930)	(1.430)	(0.919)	(1.509)	(0.992)
Age squared	0.646	1.134***	0.967	1.147**	0.042	0.106**	0.075	0.105**
	(0.628)	(0.429)	(0.666)	(0.459)	(0.073)	(0.047)	(0.077)	(0.051)
Experience	0.144**	-0.612**	0.158**	-0.670**	0.013*	-0.082**	0.014*	-0.089***
	(0.065)	(0.293)	(0.065)	(0.296)	(0.007)	(0.034)	(0.007)	(0.034)
Experience sq.	-0.009*	0.044**	-0.011**	0.048**	-0.001	0.006**	-0.001	0.006***
	(0.005)	(0.020)	(0.005)	(0.020)	(0.001)	(0.002)	(0.001)	(0.002)
Inputs:								
Home advantage	0.260***		0.259***		0.079***		0.079***	
	(0.038)		(0.038)		(0.004)		(0.004)	
Team payroll	0.406***		0.435***		0.049***		0.051***	
	(0.089)		(0.090)		(0.010)		(0.010)	
S.D. team payroll	-0.231***		-0.256***		-0.026***		-0.028***	
	(0.084)		(0.085)		(0.010)		(0.010)	
Pitcher salary	0.131***		0.133***		0.014***		0.014***	
	(0.011)		(0.011)		(0.001)		(0.001)	
Home value	0.158		0.118		0.025*		0.023	
	(0.124)		(0.125)		(0.014)		(0.014)	
Stadium age	-0.029		-0.034*		-0.002		-0.003	
	(0.018)		(0.019)		(0.002)		(0.002)	
Constant	-0.130***		-0.130***		0.461***		0.461***	
	(0.019)		(0.019)		(0.002)		(0.002)	
Team FE	yes		yes		yes		yes	
Difference in char.	no		yes		no		yes	
Observations	104,204		104,204		104,204		104,204	
R-squared	0.018		0.018		0.021		0.021	
Adj. R-squared	0.017		0.017		0.020		0.020	

Notes: Table shows suppressed variables for regression model in Table 7. *** denotes significance at 1%-level, ** significance at 5%-level and * significance at 10%-level. Standard errors are clustered at the seasonal matchup level.

Table A7: Regression Results – Match Quality and Matching on Observable Characteristics

D.V.: Match Effect	Education		Playing Level			Age	Exp.	Level Char.	Manager + Team Eff.	R-sq.	Obs.
	1 cat	2 cat	1 cat	2 cat	3 cat						
Run dif.	-0.299 (0.535)	-0.410* (0.227)						Yes	No	0.038	210
Win%	-0.150 (0.535)	-0.495** (0.227)						Yes	No	0.038	210
Run dif.	0.018 (0.114)	-0.228 (0.145)						No	Yes	0.400	210
Win%	0.013 (0.115)	-0.279* (0.153)						No	Yes	0.395	210
Run dif.			-0.292 (0.347)	-0.320 (0.294)	-0.042 (0.357)			Yes	No	0.032	210
Win%			-0.219 (0.347)	-0.330 (0.294)	0.104 (0.357)			Yes	No	0.031	210
Run dif.			-0.148 (0.197)	0.115 (0.164)	0.179 (0.174)			No	Yes	0.402	210
Win%			-0.105 (0.170)	0.151 (0.130)	0.225 (0.163)			No	Yes	0.394	210
Run dif.						-0.002 (0.073)		Yes	No	0.025	204
Win%						-0.038 (0.073)		Yes	No	0.038	204
Run dif.						0.078 (0.054)		No	Yes	0.400	204
Win%						0.059 (0.074)		No	Yes	0.385	204
Run dif.						0.121* (0.069)		Yes	No	0.020	208
Win%						0.135* (0.069)		Yes	No	0.027	208
Run dif.						-0.005 (0.053)		No	Yes	0.390	208
Win%						0.001 (0.063)		No	Yes	0.378	208

Notes: Table displays results of regressing the estimated match quality in a spell on the difference in manager personal characteristics in the pairing. Categories in education refer to no college/college attended/college graduated; categories in playing experience are not a pro player/pro player/MLB player/All-Star. The difference is expressed as the number of ordered categories matched managers differ. "Level char." is yes if the observed GM and coach characteristics are entered into the model on top of the match on the specified characteristic, no otherwise. ** denotes significance at 5%-level and * significance at 10%-level.