Wage dispersion and team performance: a theoretical model and evidence from baseball

Robert Breunig*
Australian National University

Bronwyn Garrett-Rumba
Australian National University

Mathieu Jardin
Université de Rennes I

Yvon Rocaboy
Université de Rennes I

This version: May 10, 2012

Abstract

We develop a general theoretical model of the effect of wage dispersion on team performance which nests two possibilities: wage inequality may have either negative or positive effects on team performance. A parameter which captures the marginal cost of effort, which we estimate using game-level data from Major League Baseball, determines whether wage dispersion and team performance are negatively or positively related. We find low marginal cost of effort; consequently wage disparity is negatively related to team performance. Results from game and season-level regressions also indicate a negative relationship between inequality and performance. We discuss a variety of interpretations of our results.

JEL CODES: D3,J3
KEYWORDS: wage dispersion; labor economics; sports economics; baseball; ability; effort

*Corresponding author: Research School of Economics, Australian National University, Canberra ACT, Australia; E-mail: robert.breunig@anu.edu.au.

The authors thank Deborah Cobb-Clark and José Rodrigues-Neto and seminar participants from universities around Australia for comments. The newly formed Australian Baseball League, the Canberra Cavalry and Sarge kept us going on this topic through the hot Australian summer.

1 Introduction

The enormous salaries earned by successful professional sports players frequently attract attention in the media. However, it is also well-known that less successful players, although highly paid relative to many other professionals, earn vastly less than the star players. This is certainly the case in Major League Baseball (MLB). In 2011 Alex Rodriguez was the highest paid player, earning $32,000,000, whilst many players in the first few years of their careers earn the league minimum ($414,000 in 2011). A natural question for economists is how this disparity in pay between players on a team affects the players' performances, and through them, team performance. Given a fixed wage bill, is it better to have an entire team composed of players of roughly equal wages, or to have one highly-paid star player and many less well-remunerated players? We assume below that wages reflect ability, thus one could also ask this question in terms of ability–is it better to have a team composed of players of equal ability or one star player and many 'ordinary' players? The focus of this paper is the impact of wage/ability disparity on team performance in baseball, but it should be clear that the question is of interest to labor economists in general.

There is an existing body of theoretical work considering the implications of wage disparity on productivity generally, as well as empirical work looking at the impact of inequality in wages on performance in sport. The theoretical literature is divided; Akerlof and Yellen (1991), Lazear (1989) and Levine (1991) have suggested that wage disparity may reduce productivity, whilst Lazear and Rosen (1981) and Ramaswamy and Rowthorn (1991) propose that some dispersion of wages will increase productivity. Generally, the empirical work on the impacts of pay inequality in baseball finds that wage disparity (as measured by the Gini coefficient or Herfindahl-Hirschman Index (HHI)) is associated with lower team performance.

We develop a unique theoretical model, which begins with a contest success function to derive the probability of a team winning as a function of the effort and ability levels of all players on both teams. In our model, this probability depends upon a parameter which determines whether the impact of wage disparity is positive or negative, after controlling for overall team ability. Our approach has several unique features. First, the model allows for both theoretical possibilities - negative or positive influence of wage dispersion on performance. Secondly, it allows a link between the theoretical and empirical literatures which has not been made before–the model provides a direct test of a parameter that determines which of these theoretical possibilities better matches the actual data. The model is general and can be applied to any organization where people work in teams to achieve a goal.

For our empirical analysis we use MLB data from 1985–2010 matching players' salaries to individual games. We replicate previous empirical papers which have looked at wage disparity in baseball by regressing season-level winning percentage against measures of within-team wage inequality. We extend this previous empirical work in a novel way by also estimating similar models at the level of individual games. Using maximum likelihood, we estimate the theoretical model parameter which determines whether inequality is negatively or positively related to performance. We find consistent evidence across the three approaches that inequality is associated with lower team performance, conditional on total team wage bill.

In the concluding section we offer three possible interpretations of these results: player effort is negatively affected by wage inequality and this has a negative impact on team performance; given the current salary structure and conditional on the total wage bill, teams could improve their performance by utilizing a more equal distribution of player abilities; and that the current salary distribution reflects other factors than maximizing team winning probability. Each of these three explanations is consistent with the data and they are not incompatible with one another.

In the following we briefly review the literature on contest success functions, the theoretical literature on the impacts of wage disparity generally and the empirical literature on wage disparity in baseball. We then present the theoretical model, focusing on the model parameter which determines whether within-team wage disparity increases or decreases the probability of winning. We then discuss the data used for the empirical analysis and present model results. In the final section, we provide additional discussion and conclusion.

2 Background

The theoretical model presented in the next section is based upon a contest success function, so we briefly consider work in this area. To motivate the discussion of whether wage disparity has a positive or negative impact on performance, we discuss some work from the theoretical labor economics literature on the subject. In order to place our contribution within the work on wage disparity in baseball, we discuss the approaches and results from previous work.

2.1 Contest Success Functions

The model presented in section 3 below derives the probability of a team winning as a function of the effort of all players on both teams using a contest success function (CSF). This type of function was first used in the economics literature by Tullock (1980) in the context of efficient rent seeking. The CSF was constructed from an axiomatic basis by Skaperdas (1996). Skaperdas showed that the additive CSF, which we employ, satisfies the properties of symmetry and independence of irrelevant alternatives (that is, Team A's probability of winning against Team B is unaffected by Team C). In the same paper he proved that the specific functional form used in this paper has the property of equi-proportionate changes in effort across teams leaving the probability of each team winning unchanged. These properties are desirable in our context.

2.2 Theoretical Literature

There are a large number of papers in the labor economics literature which model the impacts of wage disparity on productivity. The literature predicts that wage inequality can have positive or negative effects depending on the assumptions placed on the labor market. The model we use assumes that wage differences reflect observable ability differences and does not impose assumptions on worker responses to wage disparity. Thus, our theoretical model is silent about the mechanism which is operating to make wage inequality have either negative or positive effects on performance. The relevance of the wage disparity literature is that it yields possible explanations for the estimated value of our key model parameter and a motivation for our study.

The wage disparity literature is related to the efficiency wage literature in that both assume workers are not paid their marginal product, and that workers' effort is a function of the real wage (or expected real wage). Akerlof and Yellen (1991) gave a direct application of the efficiency wage hypothesis with a specific functional form for worker effort. Levine (1991) developed a model which also used assumptions made in the efficiency wage literature. The difference between the two areas is that the efficiency wage literature focuses on the effect of not paying workers their marginal product on employment levels, whilst the wage disparity literature focuses on the implications of wage compression or dispersion for productivity.

Lazear (1989), Levine (1991) and Akerlof and Yellen (1991) developed models in which wage disparity reduces productivity. The common theme of these models is that workers react negatively to wage disparity. Levine (1991) and Akerlof and Yellen (1991) developed models in which workers responded negatively to wage disparity due to a sense of fairness. Akerlof and Yellen (1991) modeled this directly by introducing the fair wage hypothesis. Levine (1991) proposed a similar model, but rather than workers responding directly to a fair wage, he included group cohesiveness as an increasing argument of the production function, and assumed that group cohesiveness is decreasing in wage disparity. Lazear (1989)[1] used a different approach to reach the same conclusion. He assumed wages were assigned by rank order tournament, and that workers could increase their probability of winning in two dimensions; either by increasing effort or sabotaging other workers. If workers have differing costs of sabotage, some wage compression may increase firm profits.

Lazear and Rosen (1981) and Ramaswamy and Rowthorn (1991) suggested that increased wage disparity may improve productivity, but came to this conclusion using vastly different models. Lazear and Rosen (1981) suggested that wages are determined via a rank order tournament. Workers are able to increase their probability of winning through ex-ante skill investment. It was shown that a higher wage spread gives workers a greater incentive to invest in skill, and will increase overall productivity. Ramaswamy and Rowthorn (1991) assumed that worker effort is responsive to wage. They showed that if output in some industries is more responsive to changes in effort than others, it

[1] Lazear (1989) adapts the model introduced in Lazear and Rosen (1981) to allow workers to increase their probability of winning in two dimensions. The introduction of sabotage results in the possibility of wage compression increasing productivity.

would be optimal to pay a higher wage to workers in industries where the consequences of reduced effort (referred to by the authors as damage potential) are higher. The key difference between these papers and those that hypothesize that wage disparity will have a negative impact on productivity is that Lazear and Rosen (1981) and Ramaswamy and Rowthorn (1991) developed models in which workers respond to wage disparity indirectly rather than directly.

2.3 Empirical Literature

Baseball has attracted the attention of applied economists since Rottenberg's seminal paper in 1956 (Rottenberg (1956)), which gave an economic perspective on the features of the baseball labor market, and argued in favor of free agency. A similar argument was given in Scully (1974). Since then many economic aspects of baseball have been studied, including the impact of wage distribution on team performance.

The baseball literature generally finds a negative relationship between wage disparity and team performance. Wiseman and Chatterjee (2003) conducted an empirical analysis of the impact of wage disparity on team performance, using data from 1980–2002. They measured wage inequality using Gini coefficient. Teams were divided into quartiles by Gini coefficient, and win percentage for each quartile in the periods 1980–1990, 1991-1997 and 1998–2002 was tabulated. They observed that overall salary dispersion (measured by interquartile range of Gini coefficient) had increased over the period under observation. They concluded that since the number of games won by teams in the quartile with the most inequality had decreased, whilst the number of games won by teams in the quartile with the most equal wage distribution had increased, that wage disparity must have a negative impact on team performance.

Debrock, Hendricks and Koenker (2004) came to the same conclusion using the Herfindal-Hirschman Index (HHI). They analyzed the effects of expected and residual wage disparity on team performance using data from 1985–1998. They first estimated player salaries then found that increased expected inequality resulted in decreased performance, whilst changes in residual wage disparity did not affect team performance.

Depken (2000) used data from 1980–1998 and measured wage disparity using HHI. He also concluded that wage disparity and performance are negatively related. Annual team win percentage was estimated using team salary, a linear time trend and HHI, with both fixed and random effects specifications. He found that the results were not affected by the specification chosen. This paper used similar data and measure of inequality as Debrock et al. (2004), so despite the differences in estimation technique it is not surprising that both found wage disparity has a negative impact on performance.

Avrutin and Sommers (2007) used baseball data from 2001–2005 and concluded that wage disparity does not result in decreased team performance. They regressed annual team win percentage on Gini coefficient and team payroll, and found that with the exception of 2003, neither variable was significant at the 5 per cent level. Although this fails to confirm the previous findings, it should be noted that the period under consideration was very short.

Richards and Guell (1998) also considered the relationship between team success and salary structure, however their focus was on whether teams maximise win percentage or attendance. Rather than allowing wage disparity to impact performance flexibly, it was assumed that teams maximizing win percentage would have lower salary variance. Harder (1992) also considered the impacts of wage disparity on performance, but the background of his hypotheses was from the psychology rather than economics literature, and the focus was on individual rather than team performance.

Jane (2010) considered the question of endogeneity in the relationship between salary dispersion and team performance. Using data from 1998–2007, he developed an empirical model based on that used by Depken (2000) and tested for Granger causality between total salary and performance as well as between salary dispersion and performance. He found that team performance did not Granger cause salary dispersion. The possibility of causality running in both directions is not considered in the theoretical literature, and Jane's results provide further support for the assumption of exogeneity of wage disparity which we employ in this paper.

Wage disparity has been considered in other sports. Of particular interest is Simmons and Berri (2011), which considered the impacts of wage disparity in the National Basketball Association (U.S.). They developed a theoretical model, based on Lazear's 1989 work, to motivate their empirical study. The empirical analysis followed Debrock et al. (2004) and considered the impacts of conditional and unconditional wage inequality. They found that increases in expected wage inequality, measured by Gini coefficient, increased team performance, which supports Lazear and Rosen's tournament theory hypothesis.

In the next section, a contest success function is used to derive a model of team winning probability. Following that the model is tested using MLB data, and the previous empirical studies on baseball and wage dispersion are replicated for comparability.

3 Model

In this section we develop a simple model of team cohesiveness. Our purpose is to generate hypotheses which can be tested in the data. In the model, players contribute, through effort, to the team probability of winning. Since winning has a public good aspect, and since effort is costly, players may have an incentive to shirk.

This simple model takes salaries as given and abstracts from many interesting aspects which determine baseball salaries. Players' only incentive to give effort, in the model, is to increase the probability of winning. The model is a one-period model: we abstract from effort that players might give in the goal of attaining future higher salaries, although inasmuch as these efforts simultaneously contribute to the probability of winning, one can view them as being incorporated in the player's utility function. Thus there are no explicit incentive effects of salaries and we do not model strategic interaction between owners (who determine salaries) and players.

The model is designed to investigate the effect of a particular distribution of salaries on team performance, not to determine the salary distribution. Although it would be desirable to develop a model that allows us to empirically separate cohesiveness from incentive, and hence study these strategic interactions, our inability to separately observe effort and ability makes this impossible.

The model begins by proposing a contest success function (CSF) which determines the probability of each team winning as a function of their own and the other team's effort, that is,

$$\pi^A(e^A, e^B) = \frac{e^A}{e^A + e^B} \tag{1}$$

$$\pi^B(e^A, e^B) = 1 - \pi^A(e^A, e^B) \tag{2}$$

where π^A, π^B are the respective probabilities of team A and B winning the game, and e^A, e^B is total effective effort (effort interacted with ability) of team A and team B. It is assumed that players do not behave cooperatively. Consequently, team effective effort is just the sum of individual effort multiplied by individual ability;

$$e^A = \sum_{p=1}^{n} e_p^{*A} \omega_p^A \text{ and } e^B = \sum_{p=1}^{n} e_p^{*B} \omega_p^B \tag{3}$$

We assume that e^* takes a value between zero and one. Each player has a given ability ω which he can use fully ($e^* = 1$) or not at all ($e^* = 0$) or somewhere in between. The assumption that total team effective effort is just the sum of individual effective efforts ($e_p^A = e_p^{*A}\omega_p^A$) is important for model tractability. It may be viewed as a first-order approximation to a more complicated function that includes interaction terms. For baseball, we argue that this approximation is a good one. Most outcomes in baseball are determined by the individual effort and ability of the pitcher, the hitter, the runner, or the fielder. Of course, infielders must work together to generate double plays, outfielders must communicate effectively to avoid running into each other and, perhaps most importantly, catchers and pitchers must develop strong working relationships. Nonetheless, the contribution of individual effort and ability in overall team success is much easier to measure and disentangle than in other sports which are inherently more team-oriented, such as basketball or soccer.

Total team effective effort is determined by each player choosing the best response to all other players' effective effort levels, with individuals maximizing net benefit. The net benefit function of player p on team A is given by

$$\begin{aligned} U_p^A(e_p^A) &= \frac{e^A}{e^A + e^B} - \left(e_p^{*A}\right)^\lambda \\ &= \frac{e^A}{e^A + e^B} - \left(\frac{e_p^{*A}\omega_p^A}{\omega_p^A}\right)^\lambda \end{aligned} \tag{4}$$

Players get utility from winning and pay a cost of effort. We abstract away from the utility that players might get from future expected contract benefits independent of whether their team wins or not. Player effort towards production on defense or offense will contribute both to the probability of the player's current team winning and the probability of future higher salary for the player. We assume that both effects are subsumed in the increased probability of the player's current team winning. Again, this is primarily for tractability.

Note that all players benefit equally from an increase in the probability of winning, but increasing individual effort is costly. λ is a parameter that determines the size of the effort cost. Since $e_p^{*A} \in [0,1]$ the cost of effort is decreasing in λ and the amount of effort provided will be increasing in λ.

We use player wage as a proxy for ability. For players who have achieved free agency, the free market for their talents is such that ability should be rewarded with wage. For players in the early years of their career, arbitration and market pressure make it such that these players also achieve high salaries when they show high ability. Thus, in 2011, for example, Prince Fielder of the Milwaukee Brewers, who had not yet achieved free agency nonetheless made \$15.5 million (a compromise achieved between player and team to avoid arbitration), making him one of the top paid players on the team.

Using wage, w_p^A, as a proxy for ability the net benefit equation becomes

$$U_p^A(e_p^A) = \frac{e^A}{e^A + e^B} - \left(\frac{e_p^A}{w_p^A}\right)^{\lambda} \tag{5}$$

We could allow players to be paid more than their ability but in order to generate a model that can be used to estimate λ from the data we must assume that all players are paid a salary which reflects a constant multiple above ability.[2]

Player p chooses e_p^{*A} which is equivalent to choosing e_p^A for a given wage and ability. Player p's first order condition for maximization of the net benefit function is

$$\frac{\partial U_p^A(e_p^A)}{\partial e_p^A} = \frac{e^B}{(e^A + e^B)^2} - \frac{\lambda}{w_p^A}\left(\frac{e_p^A}{w_p^A}\right)^{\lambda-1} = 0, \qquad p \in 0, \dots n \tag{6}$$

The second order condition for maximization is

$$\frac{\partial^2 U_p^A(e_p^A)}{\partial (e_p^A)^2} = \frac{-2e^B}{(e^A + e^B)^3} - \frac{\lambda(\lambda-1)}{(w_p^A)^2}\left(\frac{e_p^A}{w_p^A}\right)^{\lambda-2} < 0, \qquad p \in 0, \dots n \tag{7}$$

Clearly this will hold for all $\lambda \geq 1$. However, for $0 < \lambda < 1$, the term

$$-\frac{\lambda(\lambda-1)}{(w_p^A)^2}\left(\frac{e_p^A}{w_p^A}\right)^{\lambda-2} \tag{8}$$

[2] $w_p^A = k\omega_p^A$ with constant k across players.

is positive, so the overall sign of the second derivative is ambiguous. Let $\mu \in [0,1]$ be the value such that

$$\frac{-2e^B}{(e^A+e^B)^3} - \frac{\mu(\mu-1)}{(w_p^A)^2}\left(\frac{e_p^A}{w_p^A}\right)^{\mu-2} = 0 \tag{9}$$

For all $\lambda > \mu$, the second order condition will hold. In the derivation of the theoretical model it is assumed that $\lambda > \mu$. In the empirical application that follows, the second order conditions can be tested in terms of the observed wage data (not in terms of unobserved effort as expressed here). It is found that the second order conditions (for equations (16) and (17) below) do hold for the value of λ which maximizes the likelihood function.

The first order condition can be rearranged to give an expression for e_p^A,

$$e_p^A = (w_p^A)^{\frac{\lambda}{\lambda-1}}\left(\frac{e^B}{\lambda(e^A+e^B)^2}\right)^{\frac{1}{\lambda-1}} \qquad p=1,\ldots,n \tag{10}$$

This maximization condition holds for all players. Summing over p yields

$$e^A = \sum_{p=1}^{n}(w_p^A)^{\frac{\lambda}{\lambda-1}}\left(\frac{e^B}{\lambda(e^A+e^B)^2}\right)^{\frac{1}{\lambda-1}} \tag{11}$$

By symmetry

$$e^B = \sum_{p=1}^{n}(w_p^B)^{\frac{\lambda}{\lambda-1}}\left(\frac{e^A}{\lambda(e^A+e^B)^2}\right)^{\frac{1}{\lambda-1}} \tag{12}$$

Let $X^A = \sum_{p=1}^{n}(w_p^A)^{\frac{\lambda}{\lambda-1}}$, and define X^B equivalently. These can be rearranged and written as $X^A = (w^A)^{\frac{\lambda}{\lambda-1}}\sum_{p=1}^{n}(\alpha_p^A)^{\frac{\lambda}{\lambda-1}}$, where $w^A = \sum_{p=1}^{n} w_p^A$ and $\alpha_p^A = \frac{w_p^A}{w^A}$ for $p=1,\ldots,n$. This gives the interpretation that players' effort levels are a function of relative wage shares. Observe that

$$\frac{e^A}{e^B} = \left(\frac{X^A}{X^B}\right)^{\frac{\lambda-1}{\lambda}} \tag{13}$$

Using (11), (12) and (13), we can solve for team A's best response (i.e. the collective best response of each player on team A) to the effort level of team B, which will yield the Nash equilibrium values of team effort. Some rearranging gives

$$e^{K*} = (X^K)^{\frac{\lambda-1}{\lambda}}\left(\frac{(X^A)^{\frac{\lambda-1}{\lambda}}(X^B)^{\frac{\lambda-1}{\lambda}}}{\lambda\left((X^A)^{\frac{\lambda-1}{\lambda}}+(X^B)^{\frac{\lambda-1}{\lambda}}\right)^2}\right)^{\frac{1}{\lambda}}, \qquad K=A,B \tag{14}$$

Substituting into the CSF yields

$$\pi^A = \frac{1}{1+\left(\frac{X^B}{X^A}\right)^{\frac{\lambda-1}{\lambda}}}, \qquad \pi^B = \frac{1}{1+\left(\frac{X^A}{X^B}\right)^{\frac{\lambda-1}{\lambda}}} \tag{15}$$

or explicitly,

$$\pi^A = \frac{1}{1+\left(\frac{(w^B)^{\frac{\lambda}{\lambda-1}}\sum_{p=1}^{n}(\alpha_p^B)^{\frac{\lambda}{\lambda-1}}}{(w^A)^{\frac{\lambda}{\lambda-1}}\sum_{p=1}^{n}(\alpha_p^A)^{\frac{\lambda}{\lambda-1}}}\right)^{\frac{\lambda-1}{\lambda}}} \tag{16}$$

$$\pi^B = \frac{1}{1+\left(\frac{(w^A)^{\frac{\lambda}{\lambda-1}}\sum_{p=1}^{n}(\alpha_p^A)^{\frac{\lambda}{\lambda-1}}}{(w^B)^{\frac{\lambda}{\lambda-1}}\sum_{p=1}^{n}(\alpha_p^B)^{\frac{\lambda}{\lambda-1}}}\right)^{\frac{\lambda-1}{\lambda}}} \tag{17}$$

It can be shown that the probability of a team winning will be increasing in relative wage share. For a given total wage bill and salary distribution of the opposing team, the probability of team A winning is increasing in wage inequality if $\lambda > 1$, and the probability of winning is decreasing in wage inequality if $0 < \lambda < 1$. We can logically rule out values of $\lambda < 0$ as marginal cost of effort will be negative for these values. The impact of increasing inequality on the probability function is non-monotonic in λ, that is, the impact of higher inequality on probability of winning is increasing in λ for $0 < \lambda < 1$, and decreasing in λ for $\lambda > 1$ (see figure 1 below). The function is undefined at $\lambda = 1$, and its value as λ approaches 1 depends on whether $\lambda < 1$ or $\lambda > 1$. Table 1 presents the limiting values of the probability of winning for team A as λ approaches key limiting values where the function is discontinuous or undefined.

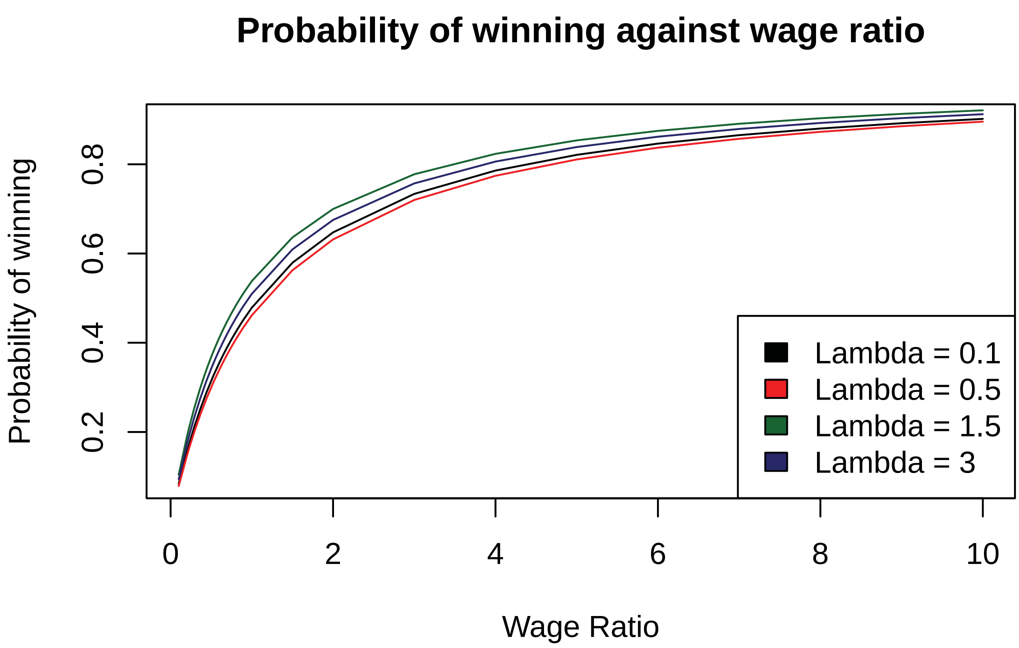
Table 1: *Limits of probability* π^A

x	0	1^-	1^+	$+\infty$
$lim_{\lambda\to x}$	$\frac{w_A\sum_{p=1}^{n}(\alpha_p^A)^{\frac{1}{n}}}{w_A\sum_{p=1}^{n}(\alpha_p^A)^{\frac{1}{n}}+w_B\sum_{p=1}^{n}(\alpha_p^B)^{\frac{1}{n}}}$	$\frac{w_A min(\alpha_p^A)}{w_A min(\alpha_p^A)+w_B min(\alpha_p^B)}$	$\frac{w_A max(\alpha_p^A)}{w_A max(\alpha_p^A)+w_B max(\alpha_p^B)}$	$\frac{w_A}{w_A+w_B}$

Figure 1 shows the impact of changing the ratio of team A's wages to team B's wages on the predicted probability of winning. In generating this figure, the wage of each player on team B has been set to $\frac{1}{9}$, and the total wage bill of team B is 1. In order to generate a non-constant probability the wage distribution of team A has been set such that one player receives $\frac{2}{9}$, 6 players receive $\frac{1}{9}$ and 2 players receive $\frac{1}{18}$ of the total wage bill. To demonstrate the impact of varying λ, we have graphed predicted probabilities for several values of λ. Figure 2 graphs the predicted probability of team A winning against λ for several different wage distributions. It illustrates that for λ between 0 and 1, the probability of winning is decreasing in inequality, whilst for $\lambda > 1$ the probability is increasing in wage disparity. However, as λ increases the gains from wage dispersion decrease. The wage distributions used are outlined in Table 2.

Using baseball data on game outcomes and player salaries for both teams, we will estimate the value of λ by maximum likelihood. The model provides the probability of winning an individual game. We can consider the salary of all players on the team for each game even if they do not appear in that game. Alternatively, we can use only salaries of players who appear in that particular game. This latter is our preferred implementation of the model. Estimation by maximum likelihood will allow us to determine whether wage disparity has a positive or negative impact on team performance.

Figure 1:



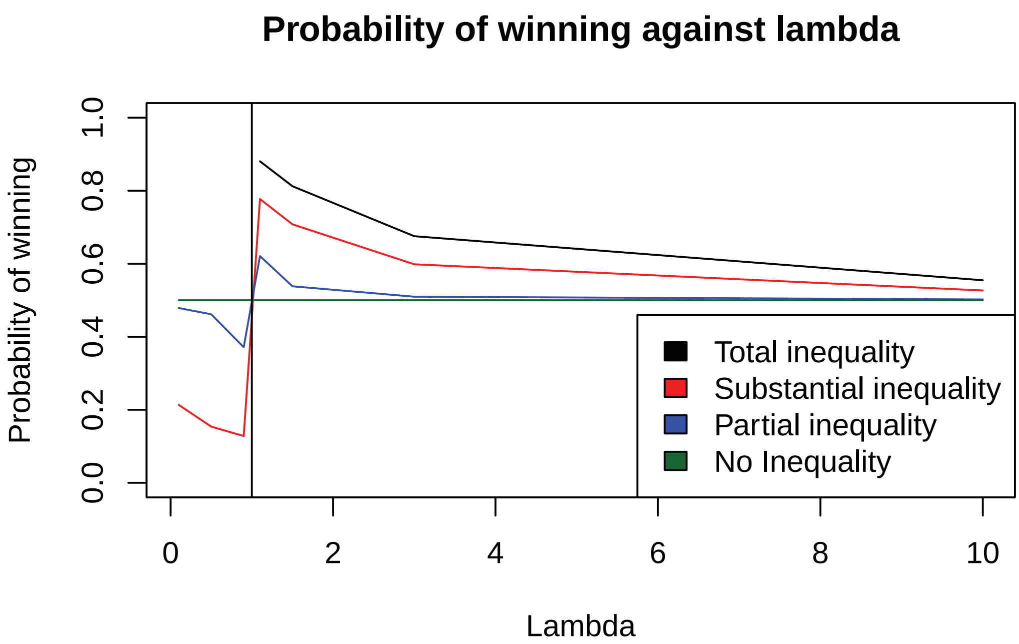


Figure 2: In the region to the left of the vertical line at $\lambda = 1$, probability of winning is decreasing in wage inequality, whilst in the region to the right it is increasing in wage inequality. For greater levels of inequality, the effect on the probability of winning of inequality is larger.

Table 2: Wage distributions

Total Inequality									
Player	1	2	3	4	5	6	7	8	9
Share of wage bill	1	0	0	0	0	0	0	0	0
Substantial inequality									
Player	1	2	3	4	5	6	7	8	9
Share of wage bill	4/9	4/9	1/36	1/36	1/36	1/36	1/36	1/36	1/36
Partial inequality									
Player	1	2	3	4	5	6	7	8	9
Share of wage bill	2/9	1/9	1/9	1/9	1/9	1/9	1/9	1/18	1/18
No inequality									
Player	1	2	3	4	5	6	7	8	9
Share of wage bill	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

4 Data

There are several aspects of baseball which make it ideal for studying the impacts of wage disparity. The abundance of data is extremely useful for gaining reliable empirical insights. Furthermore, the win/loss nature of baseball games allows for a clear measure of performance, which may be difficult to define or obtain in other areas. It can also be argued that there is a clearer relationship between player effort and performance in sport than in other occupations. Also, although clearly a team sport, in baseball players' individual performance is less related to the effort of other players than in other sports, as we argue in the preceding section. The data that we use are available online. The game level data was obtained from `http://www.retrosheet.org/`, and the salary data was obtained from `www.baseball1.com/`. The game data contained logs for 58,832 games. The salary data had 21,457 observations.

We use game and salary data from 1985–2010. Although earlier data are available the proportion of missing salaries is much higher before 1985. There is data on a total of 30 teams[3]. To estimate our preferred version of the theoretical model we require a data set with the salaries of each player appearing in that game. We use wages for the nine starting players in each game. In the National League, this is the nine starting players including the pitcher. In the American League, this is the nine starting players including the designated hitter, but omitting the pitcher. Substitutions are ignored. Again, this is a simplification, but one that avoids the inherently subjective problem of determining correct weights for players who leave a game before the end or who enter as a substitute. The importance of the contribution of a substitute is not easily measured by innings played or pitched. Key pinch-hit appearances or short relief appearances in clutch situations have disproportionate importance relative to the length of the appearance. We do estimate the model using the salaries of all team players weighted equally which includes all substitutes potential and actual and our conclusions are unaffected, as discussed in section 5 below.

Matching salary and game data was a non-trivial problem as the two data sets did not share a common player identifier. We created a new identifier for both data sets

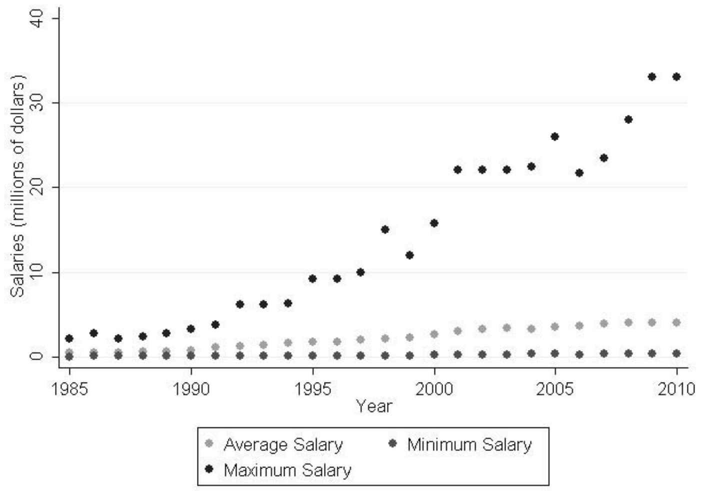
[3]Montreal Expos (1969–2004) and Washington Nationals (2005–present) are treated as the same team, as were the California Angels, Anaheim Angels and the Los Angeles Angels of Anaheim

and had to ensure uniqueness across a variety of complex circumstances. For example, in our data window there were two seasons where same-named father and son played on the same team in the same year. The existing empirical studies use annual team winning percentage in each year as the dependent variable, so the problem of matching players' salaries to games has not previously been considered.

Once this identifier was created the two data sets were merged, which resulted in almost 90 per cent of players (939,540) being matched to their salary. 11 per cent of players were left without salaries (119,436), and less than 1 per cent of salaries were not matched to players. This would occur in instances where a player does not appear in the first 9 players during a season, or when they are injured and unable to play. The existing empirical literature makes no mention of how missing salaries are dealt with. Due to the nature of the model that we are using, any game for which one or more salaries was missing would have to be discarded. Since players could play in over 100 games each year, the 10 per cent of missing salaries would result in around 80 per cent of games being discarded. For this reason we extensively researched individual player salaries and imputed some missing salaries in our data.

The imputation methods used are straightforward. League minimum salaries from 1985–2010 are available online at `http://www.baseball-reference.com/bullpen/Minimum_salary`. All players with missing salaries who appear less than 50 times (i.e. in less than 50 games in a year) were assigned the league minimum. For players who appeared more than 50 times, we attempted to find the actual salary online from a wide variety of sources–sports websites, newspaper reports, and official baseball sites. In years where players changed teams their salary was often missing even when we could discover what their new team paid. Using these methods, we found salaries for more than half of those who played over 50 times. If the salary was not available online, we imputed the salary. Players in their first year were assigned the league minimum. Players with one or two years of salaries missing during their career were assigned a midpoint. Players in the second or third year of their career with data available for subsequent years were assigned the average of the league minimum in that year and their first available salary. There are 40 players for whom the above techniques were considered questionable. These salaries were imputed using the following techniques: Most recent previous salary was used for

Figure 3: Salary distribution over time



players with at least one year of salary available, and the league minimum salary was used in all other cases. These observations have been assigned an identifier so our results can be tested for robustness. In the following empirical analysis we use three data sets. The main data set includes imputed values for all players with missing salaries. The data set which we refer to as the 'limited imputation data set' excludes imputed salaries for players who appeared less than 50 times and the 40 players for whom the imputation techniques were considered to be less reliable. The third data set excludes all players for whom salaries were missing in the data.[4]

We drop 20 games in which there was a draw (generally games that were suspended because of rain or lateness and never completed), leaving 58812 games for use in estimation. The minimum, average and maximum salaries for each year are shown in Figure 3. This gives an indication of the extent of wage disparity in baseball and its increase. Figure 4 shows the percentage of games won by the home team in each year. From this it is clear that there are distinct home and away effects. Table 3 shows average win percentage for the period 1985–2010 by team. In estimation, we need to take account of year, team and home/away effects.

[4]Full matched data available from the authors. Link to website to be provided on acceptance.

Figure 4: Home and away team winning percentages

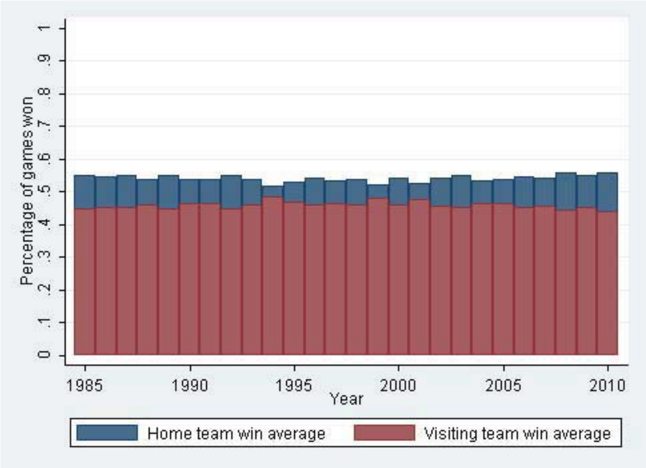


Table 3: Team winning percentages

Team	Average win percentage	Team	Average win percentage
ARI	49	MIL	47
ATL	54	MIN	50
BAL	46	NYA	57
BOS	54	NYN	52
CHA	52	OAK	53
CHN	49	PHI	50
CIN	50	PIT	45
CLE	50	SDN	48
COL	48	SEA	49
DET	47	SFN	52
FLO	48	SLN	53
HOU	52	TBA	44
KCA	46	TEX	49
LAA	52	TOR	52
LAN	52	WAS	48

5 Results

We begin by replicating descriptive regressions in the existing literature over an extended time period. Wiseman and Chatterjee (2003) and Avrutin and Sommers (2007) regress team winning percentage for each year on total wages and Gini coefficient. We regress team winning percentage for each year from 1985–2010 on average salary and Gini coefficient. The empirical model is

$$winpct_i = b_0 + b_1 avesal_i + b_2 gini_i + e_i \tag{18}$$

We have used average salary in millions of dollars (*avesal*) because the number of players appearing on a team in each year varies. In Table 4 we present the results of these regression estimates. We include models with and without imputed salaries.

Table 4: Regressions of Season Total Winning Percentage

	(1)	(2)	(3)	(4)
	winperc	winperc	winperc	winperc
avesal	.0152***	.0217***	.0107***	.0121***
	(7.35)	(9.30)	(6.90)	(7.70)
gini	-0.132**	-0.190***	-0.131***	-0.169***
	(-3.24)	(-4.59)	(-4.35)	(-5.44)
Constant	0.552***	0.584***	0.550***	0.568***
	(23.69)	(24.04)	(33.30)	(33.36)
Observations	738	738	738	738

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Model (1) follows Wiseman and Chatterjee (2003) and uses the highest 25 salaries for each team in each year to calculate the average salary and Gini coefficient. Model (2) uses the salaries for all players who appear in at least one game during the season, and includes all imputed salaries. Model (3) uses the data set with limited salary imputation and model (4) uses no imputed salaries.

As would be expected there is a significant positive relationship between average salary (acting as a proxy for wage bill and the sum of total player abilities) and team winning percentage. The coefficients of *gini* demonstrate a significant negative relationship between wage inequality and winning percentage after controlling for average team salary. The results are robust to exclusion of imputed salaries. If we use the Herfindahl-Hirschman Index (HHI) instead of the gini coefficient, we get similar results.

We also estimate, using game level data, outcomes using team wage ratio and a measure of inequality (HHI[5]). We use all games from 1985–2010, and for each game the salaries of the 9 starting players for the home and away teams. This in itself contributes to the literature in this area, as previous work has only considered annual winning percentage. This allows the salary distribution of players appearing in specific games to affect the outcome. The models are estimated using probit specifications. The basic empirical model is as follows:

$$\widehat{\Pr}(a\_wins_i) = \Phi\left(b_0 + b_1 wageratio_i + b_2 HHI_i^A + b_3 HHI_i^B\right) \tag{20}$$

where a_wins refers to the away team winning the game, $wageratio$ is the ratio of the away team's wages to the home team's wages, and HHI^A and HHI^B are the HHI for the away and home teams respectively. This model is then extended by including dummy variables for 29 home teams and 29 away teams. The results of these regressions are presented in Table 5. (Using home and away gini coefficients gives similar results.)

These results demonstrate that there is a significant positive effect of team wage ratio on the probability of winning. They also indicate a significant negative effect of own team wage disparity on the probability of winning games. The full results of these regressions are included in Appendix 1. We estimated models without HHI^B, but a likelihood ratio test rejected the null hypothesis that the coefficient of HHI^B is zero. An interesting result of the estimations is that wage disparity on the opposing team is found to increase the probability of winning. We estimated models with home/away team effects and year effects, and using a likelihood ratio test were unable to reject the null that all year coefficients were zero, but were able to reject the null that all home/away team coefficients were zero. This held in all three variants of our data set. To determine which of the estimated models is more effective in terms of predicting game outcomes, we show the percentage of correctly predicted games (Table 6).

[5]

$$HHI = \sum_{j=1}^{9} \left(\frac{w_j}{w}\right)^2 \tag{19}$$

where w_j is the salary of player j, and w is the team wage bill

Table 5: Regressions of Game-level Probability of Winning

	(1)	(2)	(3)	(4)	(5)	(6)
	A wins	A wins	A wins	A wins	A wins	A wins
A wins						
wageratio	0.0558***	0.0473***	0.0543***	0.0440***	0.0564***	0.0471***
	(13.83)	(10.60)	(7.12)	(5.20)	(5.92)	(4.40)
HHI_A	-0.483***	-0.427***	-0.452***	-0.381**	-0.583***	-0.522**
	(-6.87)	(-5.93)	(-3.47)	(-2.84)	(-3.45)	(-2.98)
HHI_B	0.557***	0.584***	0.515***	0.547***	0.530**	0.569***
	(7.99)	(8.17)	(4.00)	(4.13)	(3.19)	(3.30)
Constant	-0.195***	-0.161***	-0.178***	-0.105	-0.152**	-0.0224
	(-8.63)	(-3.48)	(-4.57)	(-1.44)	(-3.05)	(-0.24)
Home/away team effects	No	Yes	No	Yes	No	Yes
Imputed salaries	Yes	Yes	Limited	Limited	No	No
Observations	58812	58812	23200	23200	15166	15166

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These tables show a substantial improvement on the predictive power of those models which include home/away team effects. It is interesting to observe that a model which simply predicts that the home team will always win predicts 54.11% of games correctly (in the full sample of 58812 games).

Next, we estimate the value of λ from the theoretical model using maximum likelihood estimation. Given the information on which team won each game and individual player salaries, a likelihood function can be written where the likelihood contribution for games won by team A is π^A, and the likelihood contribution for games lost by team B is $\pi^B = (1 - \pi^A)$. That is,

$$L(\lambda) = \sum_{i=1}^{N}\sum_{t=1}^{T} win_{it}\pi_{it}^{A} + \sum_{i=1}^{N}\sum_{t=1}^{T}(1 - win_{it})(1 - \pi_{it}^{A}) \tag{21}$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} win_{it}\frac{1}{1 + \left(\frac{X_{it}^{B}}{X_{it}^{A}}\right)^{\frac{\lambda-1}{\lambda}}} + \sum_{i=1}^{N}\sum_{t=1}^{T}(1 - win_{it})\left(1 - \frac{1}{1 + \left(\frac{X_{it}^{B}}{X_{it}^{A}}\right)^{\frac{\lambda-1}{\lambda}}}\right) \tag{22}$$

where win_{it} is a dummy variable which takes on value one if team A wins game i at time t, and zero otherwise. X_{it}^{A} and X_{it}^{B} for game i at time t are functions of relative

Table 6: Game predictions and outcomes

Model 1 (54.89% correctly classified)

	Away win	Home win	Total
Predict away team wins	3329	2870	6199
Predict home team wins	23658	28955	52613
Total	26987	31825	58812

Model 2 (55.56% correctly classified)

	Away win	Home win	Total
Predict away team wins	6187	5335	11522
Predict home team wins	20800	26490	47290
Total	26987	31825	58812

Model 3 (54.41% correctly classified)

	Away win	Home win	Total
Predict away team wins	1222	1076	2298
Predict home team wins	9500	11402	20902
Total	10722	12478	23200

Model 4 (55.51% correctly classified)

	Away win	Home win	Total
Predict away team wins	2631	2230	4861
Predict home team wins	8091	10248	18339
Total	10722	12478	23200

Model 5 (54.29% correctly classified)

	Away win	Home win	Total
Predict away team wins	928	823	1751
Predict home team wins	6110	7305	13415
Total	7038	8128	15166

Model 6 (55.13% correctly classified)

	Away win	Home win	Total
Predict away team wins	1960	1727	3687
Predict home team wins	5078	6401	11479
Total	7038	8128	15166

wage concentration and λ as defined in Section 3.

We can augment the likelihood function to allow for home/away effects. We redefine the probability of winning, so that $\pi_{it}^{A,v} = \pi_{it}^{A} - \alpha$ is the probability of team A winning as the visiting team, and $win_{it}^{A,v}$ equals one if the visiting team (A) wins. α is chosen such that the percentage of games won by the visiting team in the sample is equal to the average predicted probability of team A winning as the visitor in the model. This gives the following moment condition:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\pi}_{it}^{A} - \hat{\alpha}) = \sum_{i=1}^{N}\sum_{t=1}^{T} win_{it}^{A,v} \tag{23}$$

Beginning by setting $\alpha = 0$, we solve for λ by maximum likelihood and then use the estimated $\hat{\lambda}$ to solve for α. The iteration is continued until convergence is achieved, as in the EM-algorithm. The final extension of the estimation involves allowing for home and away team and year effects. The probability of the visiting and home teams winning game i at time t are now redefined as follows:

$$\Pr(\text{Visiting team(A) wins}) \equiv \pi_{it}^{A,v} = \pi_{it}^{A} - \alpha_{Atv} - \alpha_{Bth} \tag{24}$$

$$\Pr(\text{Home team(B) wins}) \equiv \pi_{it}^{B,h} = \pi_{it}^{B} + \alpha_{Atv} + \alpha_{Bth} \tag{25}$$

Parameter estimates are obtained as discussed above.

The reported results are for an estimation in which λ is constrained to be positive. However, the results are not sensitive to this constraint; the largest likelihood value is still for positive values of λ. The discontinuity of the probability function at one results in different estimates of λ depending on whether the starting value is less than or greater than one. The estimation with starting value less than one gives a larger mean likelihood, so it is this result that we report. We also estimate the likelihood function for the limiting cases of Table 1, but all of these cases produce much lower likelihood values than the maximums presented in Table 7.

We report results using the data set with imputed values for all missing salaries, the data set with limited imputation and the data set with no salary imputation. We also report results allowing for home/away team and year effects. The estimated value of λ is consistently between zero and one. Table 7 also shows that allowing for home/away team effects and year effects improves the predictive success of the model.

Table 7: Estimation of λ

Data	Home/away team effects	Year effects	Coefficient(λ) (S.E).	Predictive success(%)	Sample size
Imputed salaries	No	No	0.00249 (0.00006)	53.54	58812
	Yes	No	0.122 (0.00054)	54.62	
	Yes	Yes	0.111 (0.0021)	54.74	
Limited imputed data	Yes	No	0.247 (0.0047)	54.05	23200
	Yes	Yes	0.247 (0.0047)	54.12	
No imputed data	Yes	No	0.247 (0.0065)	54.32	15166
	Yes	Yes	0.247 (0.0065)	54.23	

The implication of λ being between zero and one is that wage disparity has a negative impact on team performance. Considering the original maximization problem that players face, λ between zero and one (given $e_p^{*A} \in [0,1]$) implies a larger cost of effort than when λ is greater than one. It also implies that marginal cost of effort is decreasing. While this is somewhat surprising, sport is different from other economic activities. Professional baseball players have paid enormous fixed costs of effort to reach the major leagues. For each game, they also incur large fixed costs of warming up and marginal efforts in the last 5 minutes may be easier to make than those in the first 5 minutes. Game length is short and players may still be in the range of decreasing marginal costs of effort. This would not be the case for marathon runners, for example. λ between zero and one also implies that for constant visiting team wage inequality, increases in home team wage disparity will increase the probability of the visiting team winning. This is consistent with the results from the descriptive regressions.

Given that the baseball season is long and that many games late in the season are played between teams which are eliminated from any possibility of playing in the playoffs, we also considered estimating λ using only data from playoff games, where players might be expected to exert maximum effort. The sample size is much smaller (119 games) and we find a value of λ around 0.7, still consistent with inequality having a negative relationship with team performance. This does provide some evidence that the marginal cost of effort is higher for playoff games than non-playoff games. Instead of using game-level data, we can also calculate equation (16) summing over all players on the team instead of only summing over those players who play each individual game. Doing this results in the expressions in (16) being constant within a season for each team, although teams are still varying the team against which they are playing and win/loss outcomes vary across games. We can estimate our model (without team*year interaction effects) and cluster the standard errors to address this problem. We find very similar values for λ as those presented in Table 7.

6 Discussion and conclusion

Economists have long recognized and studied the relationship between wage disparity and individual effort when working in teams. If, for a given wage bill, the distribution of wages between workers affects the level of worker effort, then this has implications for firm decision-making. In this paper we have written down a simple model that allows for both negative and positive effects of wage inequality on team performance. We have estimated the key parameter using data from baseball and conclude that there is a negative relationship between wage dispersion and overall team performance.

The main contribution of this paper is the link between a theoretical model in which wage dispersion matters and the test of the key model parameter using wage data. We construct a unique theoretical model, built up from a contest success function, that relates individual effort to overall firm or team performance. Although we apply the model to data from Major League Baseball, the model is general. Optimal individual player effort is a function of both wage dispersion and a parameter which determines whether wage dispersion has a positive or negative effect on effort, and consequently the probability of winning. Both possibilities are theoretically acceptable within the model and the model allows these alternatives to be tested using available data. We find that the key model parameter takes a value in the range where wage dispersion is negatively related to effort.

A secondary contribution of the paper is that we extend the previous empirical literature on wage inequality in baseball by estimating the association between of wage inequality on both season-level and game-level team outcomes and examine a long data set which spans 25 years.

Our model contains several strong, simplifying assumptions which allow us to apply the theoretical model to the available data. In particular, we can not separate the effects of effort and ability as both are unobserved. Thus, our results are consistent with a variety of interpretations. If we take the model and its assumptions as given, then we would conclude that increased wage dispersion leads to lower individual effort, and consequently lower team performance. In the model, this is due to the fact that effort is costly. The key parameter value is in a range where the marginal cost of effort is high and thus players with lower ability shirk relative to those with higher ability. The benefit of giving full effort is lower for lower ability players since the decrease in team winning probability is lower for these players when they do not give full effort.

An alternative interpretation to our results is that we are uncovering a feature of the ability distribution. Consider the possibility that all players give 100 per cent effort. This may not be unreasonable as the rewards to stardom are large and when we consider players in U.S. Major League Baseball we are considering an extremely selected sample of individuals from amongst the tens of thousands of baseball players around the world trying to make the top echelon of the sport. In that case, our results can be interpreted as saying that team winning probability may be increased by having a more equal distribution of abilities within the team. A team with ten million dollars to spend may do better by hiring two five million dollar players rather than one nine million dollar player and one one million dollar player. Given the complementary nature of offensive production (home run hitters need runners on base to amass large runs batted in statistics; teams mostly score by stringing together consecutive base hits, walks, etc.) this also seems like a reasonable interpretation.

Another interpretation of our results, and one that seems absent from most of the previous empirical literature, is that the player salary structure is not chosen to optimize team winning percentage but rather to optimize the utility of the owner or the net revenues of the club. Neither of these may be maximized when team winning probability is maximized. Many owners run their clubs as a hobby and also the presence of a star player may increase gate revenues and merchandise sales well beyond the player's effect on the team's probability of winning. Data on team finances is a fiercely guarded secret, thus it would be hard to operationalize such a model.[6] However, the result that inequality is negatively related to team performance can easily be explained by such a hypothesis. Alex Rodriquez and his $32 million dollar salary contribute importantly to the inequality of the salary distribution for the New York Yankees. His value is perhaps more important to owner ego (knowing that one has bought the best, like a luxury car) and merchandise sales than to team winning probability.

The results from estimation of our theoretical model, and of the descriptive regressions which replicate what others have done previously, consistently indicate that wage disparity and team performance are negatively associated. Our results do suggest that managers attempting to maximise team performance with a given wage bill may wish to consider minimizing wage (and thus ability) dispersion. Our model is consistent with a variety of mechanisms through which this effect may be operating.

Applying the model to other types of workplace settings remains a question for future research. Innovative data which allows identification of ability, effort and wage, perhaps through matched employer-employee data which contains information on employee performance evaluations might allow for some progress to be made on this issue.

[6]Richards and Guell (1998) look at whether teams attempt to maximize attendance rather than winning, but attendance is only one part of team revenues.

References

Akerlof, G. A. and Yellen, J. L. (1991). The fair wage-effort hypothesis and unemployment, *The Quarterly Journal of Economics* **105**(2): 255–283.

Avrutin, B. M. and Sommers, P. M. (2007). Work inentives and salary distributions in major league baseball, *Atlantic Economic Journal* **35**: 509–510.

Debrock, L., Hendricks, W. and Koenker, R. (2004). Pay and performance: The impact of salary distribution on firm-level outcomes in baseball, *Journal of Sports Economics* **5**(3): 243–261.

Depken, C. A. I. (2000). Wage disparity and team productivity: evidence from major league baseball, *Economics Letters* **67**: 87–92.

Harder, J. W. (1992). Play for pay: Effects of inequity in a pay-for-performance context, *Administrative Science Quarterly* **37**(2): 321–335.

Jane, W.-J. (2010). Raising salary or redistributing it: A panel analysis of Majar League Baseball, *Economics Letters* **107**: 297–299.

Lazear, E. P. (1989). Pay equality and industrial politics, *The Journal of Political Economy* **97**(3): 561–580.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts, *Journal of Political Economy* **89**(5): 841–864.

Levine, D. I. (1991). Cohesiveness, productivity, and wage dispersion, *Journal of Economic Behaviour and Organization* **15**: 237–255.

Ramaswamy, R. and Rowthorn, R. E. (1991). Efficiency wages and wage dispersion, *Economica* **58**(232): 501–514.

Richards, D. G. and Guell, R. C. (1998). Baseball success and the structure of salaries, *Applied Economics Letters* **5**(5): 291–296.

Rottenberg, S. (1956). The baseball players' labor market, *The Journal of Political Economy* **64**(3): 242–258.

Scully, G. W. (1974). Pay and performance in major league baseball, *The American Economic Review* **64**(6): 915–930.

Simmons, R. and Berri, D. J. (2011). Mixing the princes and the paupers: Pay and performance in the national basketball association, *Labour Economics* **18**: 381–388.

Skaperdas, S. (1996). Contest success functions, *Economic Theory* **7**: 283–290.

Tullock, G. (1980). *Efficient rent seeking*, College Station: Texas A&M University, pp. 97–112.

Wiseman, F. and Chatterjee, S. (2003). Team payroll and team performance in major league baseball: 1985–2002, *Economics Bulletin* **1**(2): 1–10.

Appendix 1: Full results from regressions

Table 8: Regressions of Game-level Probability of Winning

	(1)	(2)	(3)	(4)	(5)	(6)
	A wins	A wins	A wins	A wins	A wins	A wins
A wins						
wageratio	0.0558***	0.0473***	0.0543***	0.0440***	0.0564***	0.0471***
	(13.83)	(10.60)	(7.12)	(5.20)	(5.92)	(4.40)
HHL_a	-0.483***	-0.427***	-0.452***	-0.381**	-0.583***	-0.522**
	(-6.87)	(-5.93)	(-3.47)	(-2.84)	(-3.45)	(-2.98)
HHL_b	0.557***	0.584***	0.515***	0.547***	0.530**	0.569***
	(7.99)	(8.17)	(4.00)	(4.13)	(3.19)	(3.30)
Visting team						
ARI		-0.0916		-0.0294		-0.0335
		(-1.81)		(-0.36)		(-0.33)
ATL		0.0282		0.00718		0.0277
		(0.66)		(0.11)		(0.33)
BAL		-0.106**		-0.112		-0.116
		(-2.69)		(-1.84)		(-1.47)
BOS		-0.0109		-0.0254		-0.0203
		(-0.28)		(-0.42)		(-0.26)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Continued next page

Table 8 continued

	(1)	(2)	(3)	(4)	(5)	(6)
	A wins	A wins	A wins	A wins	A wins	A wins
Visiting team						
CHA		-0.0333		-0.0460		-0.0310
		(-0.85)		(-0.78)		(-0.40)
CHN		-0.100*		-0.153*		-0.153
		(-2.34)		(-2.23)		(-1.77)
CIN		-0.0110		-0.0200		-0.0570
		(-0.26)		(-0.29)		(-0.65)
CLE		-0.0727		-0.0839		-0.0551
		(-1.85)		(-1.41)		(-0.72)
COL		-0.215***		-0.290***		-0.366***
		(-4.59)		(-3.95)		(-3.97)
DET		-0.166***		-0.222***		-0.223**
		(-4.22)		(-3.66)		(-2.87)
FLO		-0.124**		-0.178*		-0.136
		(-2.67)		(-2.45)		(-1.48)
HOU		-0.0776		-0.0673		-0.0298
		(-1.82)		(-1.01)		(-0.35)
KCA		-0.133***		-0.153*		-0.231**
		(-3.39)		(-2.54)		(-2.91)
LAA		-0.00186		-0.0454		-0.0506
		(-0.05)		(-0.75)		(-0.65)
LAN		-0.0514		-0.0973		-0.154
		(-1.20)		(-1.43)		(-1.79)
MIL		-0.105**		-0.136*		-0.0789
		(-2.63)		(-2.21)		(-1.00)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Continued next page

Table 8 continued

	(1) A wins	(2) A wins	(3) A wins	(4) A wins	(5) A wins	(6) A wins
Visiting team						
MIN		-0.0740		-0.118		-0.148
		(-1.88)		(-1.94)		(-1.87)
MON		-0.106*		-0.111		-0.118
		(-2.47)		(-1.61)		(-1.33)
NYA		0.0215		0.0148		-0.0166
		(0.54)		(0.24)		(-0.21)
NYN		-0.0547		-0.0453		-0.0394
		(-1.28)		(-0.67)		(-0.45)
OAK		0.00729		-0.0133		-0.104
		(0.19)		(-0.22)		(-1.30)
PHI		-0.0855*		-0.138*		-0.144
		(-1.99)		(-2.06)		(-1.68)
PIT		-0.195***		-0.171*		-0.130
		(-4.54)		(-2.50)		(-1.50)
SDN		-0.0800		-0.157*		-0.136
		(-1.87)		(-2.29)		(-1.54)
SEA		-0.113**		-0.124*		-0.0706
		(-2.87)		(-1.98)		(-0.87)
SFN		-0.0668		-0.0979		-0.0292
		(-1.56)		(-1.45)		(-0.34)
SLN		-0.0368		-0.0316		-0.0244
		(-0.86)		(-0.48)		(-0.29)
TBA		-0.228***		-0.255**		-0.251*
		(-4.72)		(-3.19)		(-2.40)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Continued next page

Table 8 continued

	(1)	(2)	(3)	(4)	(5)	(6)
	A wins	A wins	A wins	A wins	A wins	A wins
Visiting team						
TEX		-0.102**		-0.0904		-0.0732
		(-2.59)		(-1.45)		(-0.90)
Home team						
ARI		0.0708		0.0166		-0.0147
		(1.40)		(0.20)		(-0.15)
ATL		-0.00857		-0.00949		-0.114
		(-0.20)		(-0.14)		(-1.36)
BAL		0.177***		0.0850		0.00949
		(4.50)		(1.39)		(0.12)
BOS		-0.0657		-0.135*		-0.185*
		(-1.66)		(-2.19)		(-2.36)
CHA		0.0151		-0.0194		-0.0878
		(0.38)		(-0.32)		(-1.14)
CHN		0.137**		0.131		0.0432
		(3.20)		(1.88)		(0.50)
CIN		0.0796		0.0259		-0.0243
		(1.86)		(0.37)		(-0.28)
CLE		0.0322		0.0426		-0.0327
		(0.81)		(0.70)		(-0.42)
COL		-0.0324		0.0101		-0.0162
		(-0.69)		(0.14)		(-0.18)
DET		0.106**		0.0877		0.0697
		(2.70)		(1.43)		(0.89)
FLO		-0.00185		-0.00278		-0.0818
		(-0.04)		(-0.04)		(-0.91)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Continued next page

Table 8 continued

	(1)	(2)	(3)	(4)	(5)	(6)
	A wins	A wins	A wins	A wins	A wins	A wins
Home team						
HOU		-0.0182		-0.00491		-0.0868
		(-0.42)		(-0.07)		(-1.03)
KCA		0.132***		0.121*		0.0475
		(3.34)		(1.99)		(0.60)
LAA		0.0261		0.0224		-0.0246
		(0.67)		(0.37)		(-0.32)
LAN		0.0287		0.00684		-0.0721
		(0.67)		(0.10)		(-0.85)
MIL		0.0936*		0.0422		-0.0110
		(2.34)		(0.68)		(-0.14)
MIN		-0.0245		0.00912		-0.0218
		(-0.62)		(0.15)		(-0.28)
MON		0.0533		-0.0377		-0.138
		(1.24)		(-0.54)		(-1.56)
NYA		-0.0827*		-0.0874		-0.0937
		(-2.08)		(-1.42)		(-1.21)
NYN		0.0111		-0.0170		-0.0583
		(0.26)		(-0.25)		(-0.67)
OAK		-0.0709		-0.105		-0.120
		(-1.80)		(-1.72)		(-1.52)
PHI		0.0712		0.0646		-0.00486
		(1.66)		(0.95)		(-0.06)
PIT		0.0933*		0.0452		-0.00391
		(2.17)		(0.65)		(-0.04)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Continued next page

Table 8 continued

	(1)	(2)	(3)	(4)	(5)	(6)
	A wins	A wins	A wins	A wins	A wins	A wins
Home team						
SDN		0.0766		0.0542		-0.0141
		(1.79)		(0.79)		(-0.16)
SEA		0.0659		0.0225		0.0128
		(1.68)		(0.36)		(0.16)
SFN		-0.00226		-0.00766		-0.117
		(-0.05)		(-0.11)		(-1.35)
SLN		-0.0142		-0.0342		-0.103
		(-0.33)		(-0.51)		(-1.23)
TBA		0.0480		0.0770		0.159
		(0.99)		(0.95)		(1.46)
TEX		-0.00848		-0.00852		-0.0824
		(-0.22)		(-0.14)		(-1.04)
Constant	-0.195***	-0.161***	-0.178***	-0.105	-0.152**	-0.0224
	(-8.63)	(-3.48)	(-4.57)	(-1.44)	(-3.05)	(-0.24)
Observations	58812	58812	23200	23200	15166	15166

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 2: Extra results

Table 9: Regressions of Season Total Winning Percentage

	(1)	(2)	(3)	(4)
	winperc	winperc	winperc	winperc
avesal	.0123***	.0176***	.0091***	.0103***
	(6.19)	(7.91)	(5.84)	(6.50)
HHI	-0.374**	-0.254***	-0.310***	-0.352***
	(-4.65)	(-3.11)	(-4.72)	(-4.79)
Constant	0.521***	0.502***	0.519***	0.521***
	(51.43)	(51.18)	(56.20)	(53.38)
Observations	738	738	738	738

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table is similar to table 4 except that it uses HHI instead of gini as the inequality measure.

Table 10: Regressions of Game-level Probability of Winning

	(1) A wins	(2) A wins	(3) A wins	(4) A wins	(5) A wins	(6) A wins
A wins						
wageratio	0.0638***	0.0552***	0.0588***	0.0480***	0.0615***	0.0521***
	(16.01)	(12.44)	(7.92)	(5.77)	(6.67)	(4.94)
gini_A	-0.325***	-0.317***	-0.316***	-0.314**	-0.347***	-0.332**
	(-6.44)	(-6.11)	(-3.75)	(-3.60)	(-3.31)	(-3.05)
gini_B	0.466***	0.480***	0.333***	0.362***	0.356**	0.390***
	(9.18)	(9.19)	(4.00)	(4.21)	(3.47)	(3.63)
Constant	-0.188***	-0.131***	-0.171***	-0.073	-0.169**	-0.0148
	(-25.41)	(-3.26)	(-13.53)	(-1.18)	(-10.84)	(-0.18)
Home/away team effects	No	Yes	No	Yes	No	Yes
Imputed salaries	Yes	Yes	Limited	Limited	No	No
Observations	58812	58812	23200	23200	15166	15166

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table is similar to table 5 except that it uses gini instead of HHI as the inequality measure.

Table 11: Regressions of Game-level Probability of Winning

	(1) A wins	(2) A wins	(3) A wins	(4) A wins	(5) A wins	(6) A wins
A wins						
wageratio	0.0733***	0.0718***	0.0699***	0.0679***	0.0732***	0.0711***
	(17.61)	(17.55)	(9.45)	(9.23)	(8.02)	(7.82)
X^A	-0.0108***	-0.0692***	-0.0107***	-0.0664**	-0.0127***	-0.0794**
	(-5.21)	(-5.82)	(-3.05)	(-3.29)	(-2.90)	(-3.15)
X^B	0.0158***	0.0979***	0.0133***	0.0783***	0.0126**	0.0762***
	(7.51)	(8.12)	(3.77)	(3.90)	(2.89)	(3.05)
Constant	-0.324***	-0.570***	-0.245***	-0.336	-0.181	-0.140
	(-4.98)	(-2.90)	(-2.23)	(-1.00)	(-1.32)	(-0.33)
Value of λ	0.247	0.111	0.247	0.111	0.247	0.111
Imputed salaries	Yes	Yes	Limited	Limited	No	No
Observations	58812	58812	23200	23200	15166	15166

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table is similar to tables 5 and 10 except that it uses $X^A = \sum_{p=1}^{n} (w_p^A)^{\frac{\lambda}{\lambda-1}}$ (and X^B defined similarly) as the measure of inequality (see equation 13). We present estimates using two values of λ estimated and presented in Table 7. Here we present the results without home/away team effects. Inclusion of these effects, as in tables 5 and 10, does not change the results much.

Table 12: Game predictions and outcomes

Model 1 (55.04% correctly classified)

	Away win	Home win	Total
Predict away team wins	2723	2180	4903
Predict home team wins	24264	29645	53909
Total	26987	31825	58812

Model 2 (55.09% correctly classified)

	Away win	Home win	Total
Predict away team wins	2970	2394	5364
Predict home team wins	24017	29431	53448
Total	26987	31825	58812

Model 3 (54.63% correctly classified)

	Away win	Home win	Total
Predict away team wins	1090	895	1985
Predict home team wins	9632	11583	21215
Total	10722	12478	23200

Model 4 (54.59% correctly classified)

	Away win	Home win	Total
Predict away team wins	1139	951	2090
Predict home team wins	9583	11527	21110
Total	10722	12478	23200

Model 5 (54.55% correctly classified)

	Away win	Home win	Total
Predict away team wins	826	680	1506
Predict home team wins	6212	7448	13660
Total	7038	8128	15166

Model 6 (54.52% correctly classified)

	Away win	Home win	Total
Predict away team wins	873	732	1605
Predict home team wins	6165	7396	13561
Total	7038	8128	15166

The models referred to are those of Table 11.