The effect of wage dispersion on performance in the NBA

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Industrial Dynamics and Strategy

Supervisor: Thomas Peeters

Second reader: Francesco Principe

Name: Jason Ching

Student number: 367843

E-mail address: jasonching@live.com

Contents

1. Introduction

In a professional sport setting there are numerous discussions regarding the large disparity of salaries between players. Sports teams have a choice in setting the level of wage disparity between their players, having either a high or low level of inequality. They can have a very unequal distribution of salaries by having one super star player with very weak supporting cast of teammates. On the other hand, teams can choose to hire players with perceived equal skill and thus have a fair pay structure for all players. What is the optimal salary structure for a team to increases performance? Is there value in paying high wages for exceptionally skilled players within a sports context? The O-ring theory of production highlights the importance of this relationship between high skilled workers, wages and production (Kremer, 1993). Many firms have production processes where the cooperation between several workers is needed to deliver a finished product, a firm is only as strong as its weakest link. A weak link in the chain can have significant impacts on the final product. A firm can have exceptional product design and marketing, however if the manufacturing is not up to the same quality as the other departments the final product will result in a failure. Therefore, the O-ring theory states that due to this interdependence, high skills workers will be clustered together and small difference in skills between workers will result in large difference in wages (Kremer, 1993). This process has led to several studies examining the effect of salary disparity in a sports setting, with Major League Baseball receiving the most attention, having extensive studies indicate a negative relationship between performance and wage dispersion (e.g. Bloom, 1999; Depken, 2000; Frick et al, 2003; Jewell and Molina, 2004; DeBrock et al, 2004).

The National Basketball Association (NBA) is an interesting case as previous attempts to shed light on this relationship have conflicting results. In some previous studies there was evidence that salary dispersion had no effect on team performance (e.g. Berri and Jewell, 2004; Katayama and Nuch, 2011). However, other studies by Frick et al. (2003) and Simmons and Berri (2011) find a positive effect on team performance for increasing levels of salary inequality. The aim of this paper is to further investigate this link between payroll dispersion and performance for the NBA. This paper uses team payroll data from the NBA from 2008 to 2018. The first analysis will focus on the relationship between payroll distribution and team performance, to determine if the structure of salary distribution does affect the performance of a team. Secondly, I distinguish this paper from those in the literature by also looking at the predictive power of this model and compare it against several benchmark forecasting models within the sport forecasting literature. I then create predictions for NBA games to evaluate the monetary returns when using several betting strategies. This reveals how wage dispersion affects production at the individual team level and the monetary gains that can be generated when incorporating salary dispersion into forecasting models.

Betting markets share many similarities with financial markets that make them useful in analyzing professional sports. Individuals in both markets are heterogeneous, profit-maximizing investors that have access to different levels of information to deal with diminishing risk until the trading deadline (Levitt, 2004). Additionally, in both markets extremely large amounts of money are staked in a zero-sum game between two individuals. According to Weissman (2014) the global sports betting market is currently estimated to be valued at 60 billion USD. However, a large portion of sports betting is not reported as the illegal sports betting has a reported value of 400 billion USD. These similarities have made the use of betting odds to be extremely useful in comparing forecasting accuracy in similar (Forrest et al., 2005).

This paper finds evidence of a link between wage dispersion and performance within the NBA. Specifically, an inverse U-shape between these two variables, indicating that beyond a certain point it does not pay to increase the wage disparity between players. In a forecasting setting, I find that including wage dispersion is an improvement over a naive model but not significantly different from a model with only variables for team payroll and home advantage indicator. In addition, the wage dispersion model does not outperform the implied probabilities from betting odds. Lastly, when evaluating the betting returns, a model including wage dispersion does outperform a naïve model and a model using only home advantage and team payroll. These interconnected results demonstrate that wage dispersion does affect performance in the NBA, but the magnitude of the effect does not have a large impact.

The following section describes the theoretical framework regarding salary dispersion and team performance and a summary of previous results. Section 3 describes the data set used in my analysis and the method and models used to forecast NBA games. Section 4 evaluates the performance of my prediction model and its forecasting performance. Additionally, I also present the returns from several different betting strategies. Section 5 provides a conclusion and possible areas for further research.

2. Related literature

2.1 Theoretical framework

There are two competing theories regarding the effect of salary dispersion on team productivity (Franck & Nüesch, 2011). Firstly, the hierarchical pay hypothesis which proposes a positive link between pay and performance when wages are dispersed. Players will often asses the value of their salary through a process of social comparison (Festinger, 1954). Milgrom and Roberts (1992) find that large wage differences are inevitable to attract and keep talented players. Furthermore, the creation of a positive link between performance and pay increases the returns for high performing players, inducing greater performance in the future. Ramaswamy and Rowthorn (1991) propose a similar model where wage dispersion also has a positive impact on productivity. They imply that differentiated players introduce different levels of 'damage potentials' to team productivity as they could interfere and harm productivity due to adverse feelings created by high levels of salary dispersion. Thus, some players require greater wages to mitigate their motivation to inflict damage to team productivity. Lazear and Rosen (1981) introduce tournament theory to explain the benefit of a hierarchical pay structure to improve team productivity. Under tournament theory, workers are given a rank based on their relative performance for the firm and are rewarded based on their rank. This promotes optimal effort from workers due to receiving relatively greater rewards when promoted. Furthermore, a key aspect to increase productivity is to have a high spread in rewards, a greater difference in rewards leads to higher worker effort.

The wage compression hypothesis proposes that team performance reacts negatively as wage dispersion increase. An increase in pay inequality reduces satisfaction over both pay and job and creates feelings of envy, unfairness and resentment to the detriment of team performance relative to a more equal pay structure (Simmons and Berri, 2011). Martin (1981) shows that workers experience feelings of deprivation under an unequal pay structure. The level of social discontent within a team is mostly determined by the relative comparison of an individual player relative to the social and economic position of other team mates (Franck and Nuesch, 2011). Levine (1991) proposes that an equal pay structure produces and fosters group cohesion in teams, which would lead to increases in team productivity.

2.2 Previous results

Studies using the NBA to identify the impact of wage dispersion on performance have found inconsistent results. Frick et al. (2003) looked at the effect of Gini coefficients of player salary on seasonal winning percentages for the NBA, National Football League and the National Hockey League. They were only able to find a positive effect for wage dispersion in the NBA, a higher level of salary dispersion was linked with higher levels of performance. However, Berri and Jewell (2004) find that salary dispersion has no significant impact on team performance. Their research looked at a different period within the NBA, looking at six consecutive seasons and found no evidence of wage dispersion affecting team performance. They suggest that NBA players are not sensitive to salary differentials during this time and thus find no significant results. Further research by Katayama and Nuch (2011) expand on these two previous studies by also looking at game level data instead of just seasonal measure of game performance. They examine the NBA for five consecutive seasons and find that there is no significant relationship between wage dispersion and team performance.

The empirical literature and the previous results clearly show that there are competing theories and conflicting results. The previous results have either provided evidence for both the hierarchical pay hypothesis and the wage compression hypothesis or they have evidence stating that wage dispersion has no impact on team production. This paper will expand on this previous literature and shed further light on the effect of wage dispersion.

3. Estimation Approach

3.1 Data

I investigate the link between team performance and team payroll distribution from a dataset containing all games for the NBA between October 2008 and June 2018, consisting of nine seasons with pre-season, regular and playoff games. The data comes from basketball-reference.com, a popular basketball statistics website. The website contains information on all games played in the NBA season from 2008 to 2018 with data on team payroll for each season. Additionally, it contains data at the individual game level allowing us to see the exact score and whether each team played at home throughout the seasons. Table 1 contains summary statistics on the dataset. There are 27,434 observations over the 9-year period, for a total of 13,717 games. The data contains two observations for every game played during the season, one observation from each team's perspective. This was done so that the control variable for home advantage would be balanced across all teams. The literature has shown that home advantage has a significant impact on game outcomes (Garicano, Palacios-Huerta, & Prendergast, 2005).

Table 1 contains a summary of all the variables in the dataset. Team performance is measured using two separate variables, a discrete measure of a win or loss and the score difference between the two teams. The *Win* variable takes value 1 if the team won the game or 0 if the team lost the game. The *score margin* variable is the difference in points scored between the two teams. The score margin is an important measure of performance as it contains more information regarding the outcome of a game. A team that wins by a 50-point margin has significantly outperformed the opposing team, but this information would not be captured in the performance variable of *Win*. Following Franck and Nuesch (2011) I use two alternative measures of pay dispersion as explanatory variables. Firstly, I use the *Gini coefficient*, a measure of inequality frequently used for income distributions. This measure takes a value between 1 and 0, with a value of 1 representing maximum inequality and 0 representing perfect equality. Secondly, I use the *coefficient of variation*, another measure of dispersion calculated as the ratio of the standard deviation to the mean of a sample. These variables are calculated at the individual team level for each NBA season. The *home* advantage indicator is a dummy variable taking the value 1 when a team plays at their home stadium and otherwise 0. The variable *Team payroll* is the total salary expenditure for all players on a team for a given season.

Table 1: Summary statistics for forecasting and control variables

Variables	Obs.	Mean	Std. Dev.	Min	Max
Win	27 434	0.50	0.50	0	1
Score margin	27 434	0	13.71	-61	61
Home	27 434	0.50	0.50	0	1
Team payroll	27 434	77 574 109	18 300 000	45 268 465	137 494 845
Gini coefficient	27 434	0.53	0.08	0.22	0.74
Coefficient of variation	27 434	1.09	0.22	0.41	1.74
Betting odds	27 434	2.63	2.19	1.01	28.13

Note: Team payroll is reported in USD.

3.2 Baseline estimation

I introduce two regression models for linking team payroll and performance for the NBA. A probit model using game outcome as the dependent variable and an ordinary least squares (OLS) model using score margin as the dependent variable (Peeters, 2018; Frank and Nuesch, 2011). For the first model, a probit with the dependent variable as the variable *Win*, y_{ijt} that describes the result of the game between team *i* and *j* as either a win or loss at time *t*. I relate this outcome to the explanatory variables of team payroll and home. The linear form of the model can be specified as

$$y_{ijt} = \beta_h h_{ijt} + \beta_p(p_{it} - p_{jt}) + \beta_G(D_{it} - D_{jt}) + \beta_G(D^2{}_{it} - D^2{}_{jt}) + \epsilon_{ijt}$$

where h_{ijt} the home is advantage indicator, p_{it} and p_{jt} are the logged team payrolls, D_{it} and D_{jt} are measures for payroll distribution, and ϵ_{ijt} represents the error term. I expect to find positive coefficients for the home and team payroll variables, as a team with home advantage and a higher average payroll should have a higher probability of winning the game. I expect to find a U-shaped relationship for the payroll distribution variables, as it would support the hierarchical pay hypothesis and the wage compression hypothesis. Team performance would be maximized at both high and low levels of inequality, similar to the findings by Frank and Nuesch (2011). I assume that both teams contribute to the outcome of the game and therefore specifically focus on the difference between team i and j for the variables p and D. The squared terms for the distribution measures are calculated by first taking the square of each term and then taking the difference. Furthermore, this controls for seasonal differences in average payroll levels, since it is more important how much more a team spends on payroll compared to another team than the absolute value of their payroll. Using similar notation as above, the second model uses the score margin as the dependent variable in an OLS model. This model will then capture the extra variation that can be explained through large or small differences in points scored. As mentioned previously, teams that wins by large margins compared to teams with narrow wins clearly demonstrates a higher difference in perceived skill than what the win variable captures.

3.3 Forecasting approach

Using the regression models, I predict the outcome for a test sample of NBA games and benchmark the forecasting performance by comparing it to 3 competing prediction models. A model with only the home advantage indicator and team payroll, a naïve prediction model and averaged decimal betting odds. Under a naïve prediction model, I assume that the home advantage variables indicates the winner of a game, this would represent a lower bound of predictive performance. Betting odds are used as they are a popular benchmark for comparing forecasting accuracy in the sports forecasting literature (Forrest, Goddard, &

Simmons, 2005); (Franck, Verbeek, & Nuesch, 2010). Bookmakers can face serious financial consequences when they misprice bets and can incorporate all available data leading up the game day to price their bets, therefore they most likely represent an upper bound for forecasting performance (Peeters, 2018). The decimal betting odds were collected from oddsportal.com and represent an average of decimal odds across different bookmakers.

The inverse of bookmaker odds can be interpreted as the probability of a given team winning. However, using data from odd-setters require a transformation as they still contain the "over-round" that bookmakers include in their decimal odds. I calculate the implied probabilities by taking the inverse of the odd and subtracting the over-round. I first calculate the over round by

$$Over_{ijt} = \frac{1}{odd(win)_{ijt}} + \frac{1}{odd(loss)_{ijt}} - 1$$

I can then calculate the implied probabilities of winning using

$$Prob_{(win_{ijt})} = \frac{1}{odd(win_{ijt})} \frac{1}{1 + Over_{\ ijt}}$$

For example, on the 31st of May 2018, the Cleveland Cavaliers played the Golden State Warriors in the first game of the NBA finals. In the table below, I list the decimal betting odds for this game and the calculated implied probabilities.

Table 2: Example of decimal betting odds

Team	Odds	Prob. with Over	Over	Implied Prob.
Golden State Warriors	1.11	0.90	0.04	0.87
Cleveland Cavaliers	7.2	0.14	0.04	0.13

Based on these decimal betting odds the probability of the Golden State Warriors and Cleveland Cavaliers winning are 0.90 and 0.14, respectively. However, these probabilities sum to a value greater than one as the over round of 0.04 from the bookmaker is still included. Using the equations above, the real winning probabilities are 0.87 for the Golden State Warriors and 0.13 for the Cleveland Cavaliers. This allows for a direct comparison between forecasted probabilities from a payroll model to the decimal betting odds.

4 Results

4.1 Baseline estimation results

In table 3, I present the estimation results using the full data set. This provides an overall view of the fit of the regression models and in which direction the explanatory variables affect the dependent variables. Table 3 shows 5 different regression models using the full data set, the first 3 columns are probit regressions using game outcome as the dependent variable, here I can see the difference when including a measure for payroll distribution. The last two columns are the OLS regression models, first with Gini Index and then with the coefficient of variation.

Across all the probit models, teams with a higher payroll relative to their opponent have higher probabilities of winning games, as the variable *Wage* is positive and significant in all models. The variable *Home* is also positive and significant, indicating that teams playing at home have a higher probability of winning games compared to teams playing away. Looking at only the linear terms, including a measure for salary distribution adds further explanatory power to the model as the variables are both significant at the 1% level. A higher Gini coefficient relative to the opposing team increases the probability of winning a game. A higher coefficient of variation relative to the opponent team also increases the probability of winning a game.

These results are in line with the findings by Frick et al. (2003) and Simmons and Berri (2011), where wage dispersion has a positive linear effect on team performance. However, the squared terms for the Gini coefficient and coefficient of variation are both negative and significant at the 1% level, indicating evidence of an inverse U-shaped relationship with game outcome. Teams underperform when having an extremely fair pay structure or a highly unequal pay structure, in this case it pays to have a salary structure in the middle, between the two extremes. Team performance is maximized with a Gini coefficient of 0.71 and a coefficient of variation of 0.89. However, for the Gini coefficient, this value is close to the observed maximum value of 0.74 for all the observations. In reality, it could be very difficult to obtain this level of salary distribution as the declining section is not as relevant in the sample. A possible explanation is highlighted by Coates, Frick and Jewell (2014), the cooperative nature of basketball can require substantial coordination among all the players on the court. A superstar player will still need to rely on his teammates to perform as individual plays do not often determine team success. Therefore, increasingly high levels of salary inequality can be detrimental to the overall success of the team.

Table 3: Estimation results for full dataset

Outcome	Game result			Score margin	
Wage	0.486*	0.969*	0.958*	10.036*	9.893*
	(0.015)	(0.041)	(0.041)	(0.420)	(0.416)
Home	1.089*	0.487*	0.488*	5.788*	5.789*
	(0.039)	(0.015)	(0.015)	(0.158)	(0.158)
Gini		4.203*		41.911*	
		(0.601)		(6.136)	
$Gini^2$		-3.527*		-34.685*	
		(0.574)		(5.874)	
CV			2.036*		21.606*
			(0.189)		(1.929)
CV^2			-0.844*		-8.950*
			(0.083)		(0.845)
Constant	-0.243*	-0.244*	-0.244*	-2.894*	-2.895*
	(0.011)	(0.011)	(0.011)	(0.112)	(0.112)
Model	Probit			OLS	
Obs.	27 434	27 434	27 434	27 434	27 434
Pseudo R^2	0.047	0.050	0.049	0.077	0.077
AIC	1.322	1.318	1.319	7.994	7.996
BIC	-246 395.2	-246 500.3	-246 459.1	-61 775.3	-61 727.3
Log-Likelihood	-18 281.7	-18 218.9	-18 239.5	-110 581.4	-110 605.4

Note: Standard errors are reported in parentheses. The last column reports values with the coefficient of variation. The second to last column reports values with the Gini Index. * Significance: $p < 0.01$.

The OLS models have similar results to the probit models across the explanatory variables. The variable for team payroll and home advantage are both positive and significant. Furthermore, under these models there is also evidence for an inverse U-relationship between salary dispersion and game performance. Score margin is maximized with a Gini coefficient of 0.60 and a coefficient of variation of 1.21. These values are closer to the observed mean for both the Gini coefficient and coefficient of variation, indicating that both the declining and rising part are relevant in the observations.

In terms of goodness of fit, I use four different measures, pseudo R^2, Akaike information criterion (AIC), Bayesian information criterion (BIC) and log-likelihood. The pseudo R^2 shows to what extent the model is an improvement over one with just a constant term, a value closer to 1 means a better fit given the data. The pseudo R^2 indicates that the first model performs the worst, when no distributional measures are included. The probit model with Gini coefficient is slightly better than the coefficient of variation probit model according to this measure. The AIC and BIC parameters judge the quality of a model by having a trade-off between the explanatory power of a model and the number of coefficients. Lower AIC and BIC values show a better fit when comparing across models. For the probit models, adding the distributional models led to a lower AIC and BIC values, between these two models, the model using Gini coefficient has a slightly better fit than the model with coefficient of variation. This also holds true for the OLS models, where the model with Gini coefficient has a slightly better fit. The log-likelihood parameter shows the probability that the observed data is true, given the model. A higher value represents a better fit for the model using the observed data. This measure of fit has similar results as the previous measures, with the model using Gini coefficient performing slightly better than their counterparts. Therefore, in the remainder of this paper I report results using the probit model with the Gini coefficient as the measure for payroll distribution.

4.2 Forecasting performance

In forecasting I can only use all the available information before the start of an event when predicting the result for that event. Therefore, I estimate the regression models on an initial training sample of data using a rolling estimation algorithm. The initial window is 40% of the full data set to train the model, I then re-estimate the model for each game day using only the observations that came before it. I then use the new estimates to forecast the results for all games played on that game day.

Following Peeters (2018), I use Brier scores, pseudo-likelihood statistics and success ratios to compare the probit model against the two alternative prediction models. The Brier score is a quadratic loss function comparing the predicted probabilities with the observed outcomes. Brier scores can be evaluated across different models, as a lower Brier score indicates better forecasting performance. Furthermore, I examine the significance of differences in prediction accuracy by performing pairwise t-tests at the level of individual observations. This would indicate if Brier scores from two different models are significantly different. The pseudo-likelihood is a measure for the amount of information needed to communicate the observed outcomes of games given the true probabilities for each result. The average value for this statistic across the given sample can be compared across models. A higher pseudo-likelihood value indicates superior forecasting performance. Lastly, I compare the success ratio across my two forecasting models. The success ratio measures the correct number of discrete game outcome predictions. For each game, I assign a discrete predicted outcome (win or loss) based on the higher probability of winning or losing. If the actual outcome of the game is then equal to the predicted discrete outcome it is labeled as a success. The success ratio is then the total number of successes divided by the number of total games in the sample. This allows for a clear indicator of correct predictions for each forecasting model. I also include a 'Payroll' model that includes all the explanatory variables except the payroll distribution, this was done to compare the improvement in forecasting with and without payroll distribution. Additionally, I test if the initial window size of my model will influence the forecasting performance by running the model using an initial window size of the first 20 game days.

Table 4: Forecasting performance of prediction models

Sample	Games	Forecaster	Brier Score	Pseudo-likelihood	Success ratio
Last 60% of games	16 617	Full	0.2363	0.6655	0.5978
		Betting odds	0.2030*	0.5902	0.6844
		Payroll	0.2365	0.6657	0.5978
		Naïve	0.4124*	-	0.5876
All games after 20 match days	27 380	Full	0.2336	0.6605	0.6082
		Betting odds	0.2011*	0.5862	0.6874
		Payroll	0.2342**	0.6612	0.6051
		Naïve	0.4059*	-	0.5940

Note: The brier scores are averaged at the game level and * denotes significance at $p < 0.01$, ** denotes significance at the $p < 0.05$ for the pairwise t-tests against the brier score values for the full forecasting model.

Table 4 presents the results for the different models and samples of the data set. As mentioned previously, the models need 'training' data to increase accuracy of the model. Therefore, the forecasted results are only calculated after the initial training data. The first panel represents 16,617 games according to 4 different prediction models. The betting odds model performs the best as it has the lowest brier score and pseudo-likelihood values and the highest success ratio, being able to predict 68.44% of games correctly. Comparing the full model to the payroll model there are mixed results, the brier score for the full model is slightly lower than the brier score for the payroll model, indicating that adding a distribution measure has improved the forecasting performance. However, the pair wise t-tests between the two brier scores indicates that the payroll brier score is not significantly different from the brier score of the full model. Thus, the addition of a distributional measure does not improve the forecasting performance over a model with only home advantage and team payroll. This is also indicated by the success ratio, as the two models make the exact same number of predictions correctly. Comparing the full and payroll models to the naïve model, there is a significant improvement in Brier scores and higher success ratios. The naïve model does perform the worst out of all the prediction models and does represent the lower bound for predictive performance across the models. There are no pseudo-likelihood values for the naïve model as they use discrete game outcomes as their prediction, resulting in undefined values.

The second panel of Table 4 is a robustness check of the results by choosing the smallest possible learning period to see if the choice of learning period influences the forecasting accuracy. I can then evaluate the forecasting performance when looking at the largest possible number of games. In this sub sample the results remain largely the same as the first panel, with the betting odds performing the best and the naïve model performing the worst. However, one difference is that the brier score for the payroll model is significantly different from the full model at $p<0.05$, indicating that in a larger sample adding distributional measures does improve the forecasting performance over the payroll model. This could be due to more observations increasing the likelihood that a result is significant, so a small effect may become more significant with more data. Overall the different sample size doesn't appear to show a very large difference between the models, as such the results do not seem to be affected by the choice of the testing period.

4.3 Returns from betting

Lastly, I calculate the returns for several betting strategies based on the Payroll model predictions. Following Hvattum and Arntzen (2010) and Peeters (2018), I use three betting strategies, unit bet, unit win, and Kelly bet. The betting returns for my data set are calculated using the raw betting odds, they will still include the over round from the bookmaker. Therefore, in reality the returns on naïve betting strategies will be negative and a successful strategy may result in a less negative return. All three betting strategies have a bettor place a bet when the expected return for the strategy is positive. Using similar notation to previous sections, the expected return for a game between team i and j at time t can be calculated as follows

$$E(R_{yijt}) = S_x * odd(y_{ijt}) * f_{ijt-1}(y_{ijt}) - S_x$$

where S_x is the stake for the betting strategy used, $odd(y_{ijt})$ is the betting odds for the given outcome, and $f_{ijt-1}(y_{ijt})$ is the forecasted probability of outcome y_{ijt} using all available information leading up to the current game day.

A unit bet strategy has a bettor place a one unit stake on every bet with a positive expected return. This will give a return of one multiplied by the betting odds on correctly predicted bets or a loss of one unit on incorrect predictions. As the name entails, the stake is

$$S_{unit\ bet} = 1$$

An issue with this strategy is that bets on very probable outcomes yield very low returns as the stake remains at one across all bets regardless of the odds on the game. A unit win strategy can account for this by setting the stake on every bet to yield a return of one unit when it predicts correctly. Under this strategy a bettor places a stake on each bet with a positive expected return but takes into account the betting odds in their stake for each bet. This stake can be calculated as

$$S_{unit\ win} = \frac{1}{odd(y_{ijt}) - 1}$$

However, an issue with this strategy is that many bets require placing very large stakes in order to yield a one unit return. The last betting strategy I use is a Kelly bet, the bettor once again places a stake on bets with a positive expected return, but the stake is designed to maximize the long term wealth of the bettor. The advantage of this strategy is that it not only takes into consideration how much should be staked but also when. The Kelly stake can be calculated as

$$S_{Kelly} = \frac{f_{ijt-1}(y_{ijt})(odd(y_{ijt}) - 1) - (1 - f_{ijt-1}(y_{ijt}))}{odd(y_{ijt}) - 1}$$

In the table below, the betting returns are summarized using the same samples as in table 4. I present the mean return across all bets that had a positive expected return under each betting strategy.

Table 5: Returns from betting

Sample	Games	Model	Unit bet		Unit win		Kelly bet		No. of bets
			Mean stake	Mean return	Mean stake	Mean return	Mean stake	Mean return	
Last 60% of games	16 617	Full	1.00	-4.74%	0.64	-2.76%	0.22	-0.77%	7 690
		Payroll	1.00	-4.97%	2.28	-3.06%	0.22	-0.83%	7 673
		Naive	1.00	-4.19%	3.35	-9.08%	0.51	-1.87%	13 717[a]
After 20 game days	27 380	Full	1.00	-5.47%	0.72	-2.84%	0.22	-0.90%	12 573
		Payroll	1.00	-6.22%	1.66	-3.55%	0.22	-1.10%	12 573
		Naive	1.00	-4.19%	3.45	-9.68%	0.51	-2.03%	13 717[b]

Note: [a] Number of bets under the unit bet strategy, 8 308 and 10 437 for unit win and Kelly bet respectively. [b] Number of bets under unit bet strategy, 13 704 and 17 145 for unit win and Kelly bet respectively.

Across all betting strategies the mean returns are negative, however comparing across the three strategies the returns do improve according to betting strategy. The unit win, and Kelly bet consistently outperform the unit bet strategy with the exception of the naïve model. For the full and payroll model it is clearly better to adjust the stakes according to the odds. Comparing between the models, the full model does outperform the payroll model across all three betting strategies, indicating that including the Gini index into the model does have monetary benefit when betting. The mean return increases by 0.23%, 0.30% and 0.06% for the unit bet, unit win, and Kelly bet respectively. Also, under the unit win strategy the full model is able to generate a lower return while having a much lower capital exposure as the average stake is 0.64 to the average stake of 2.28 for the payroll model while using a unit win strategy. The last panel once again shows the same analysis but for all games after the first 20 days. The results are very similar to the first panel, with full model outperforming the payroll model. Also, under these two models adjusting the stake to the odds improves the average return.

5. Conclusion

The NBA is an interesting area of research when dealing with salary dispersion, as many studies report conflicting results regarding its effect on team performance. The aim of this paper was to further investigate this link between wage dispersion and performance in the NBA. This paper has taken three distinct methods in evaluating the effect of wage dispersion on team performance in the NBA. The first method, which follows previous attempts in the literature by looking at the regression results from a large sample of NBA games, finds evidence that wage dispersion has a positive effect on team performance within the NBA. Although, there are diminishing returns to increasingly high levels of wage dispersion, as there is evidence of an inverse U-relationship between salary dispersion and team performance. I then evaluate the forecasting performance of this model against other competing models. I find that wage dispersion does not significantly improve the forecasting performance over a model using only a home advantage indicator and team payroll. Although, I do find that including wage dispersion does add predictive power over a naïve model. Lastly, when evaluating the betting returns I find that my model using salary dispersion produces very small negative returns. These negative returns are smaller than what would be yielded under a naïve model, therefore including a measure for salary dispersion does increases the returns when betting on NBA games.

Together these results suggest that wage dispersion does impact game performance as suggested by the existing literature. However, the magnitude of the effect has a relatively small impact on overall team performance. In forecasting NBA games, the most important sources of information are home advantage indicator and the amount of money invested in team payroll.

This study has several extensions for future research to expand on the relationship between salary dispersion and team performance. First, it is plausible that if wage dispersion affects team performance, it can also affect how team output is produced. Lazear (1989) predicted that members of high inequality teams produce wins in a less cooperative fashion, relying more on a selfish method of production relative to teams with an equal salary structure. A possible method to identify this production method is shown by Franck and Nuesch (2011) for soccer in Germany. They identify cooperative plays and individualistic plays by looking at the number of passes and dribbles in a game between players and identify the effect of wage dispersion on these variables. Lastly, an area for future research is to generalize the results of this study, it remains to be seen if these results carry over to other sports.

6. References

Allison, P. D. (1978). Measures of inequality. *American Sociological Review, 43*, 865–880.

Berri, D. J., & Jewell, R. T. (2004). Wage Inequality and Firm Performance: Professional Basketball's Natural Experiment. *Atlantic Economic Journal, 32*, 130-130.

Bloom, M. (1999). The performance effects of pay distribution on individuals and organizations. *Academy of Management Journal, 42*, 25-40.

Butler, M. R. (1995). Competitive Balance in Major League Baseball. *The American Economist, 39(2)*, 46-52.

Coates, D., Frick, B., & Jewell, T. (2014). Superstar Salaries and Soccer Success: The Impact of Designated Players in Major League Soccer. *Journal of Sports Economics, 17(7)*, 716-735.

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

Depken, C. A. (2000). Salary disparity and team productivity: Evidence from major league baseball. *Economics Letters, 67*, 87-92.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Forman, S. (2019). *Basketball Reference* . Retrieved from https://www.basketball-reference.com/

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting, 21*, 551-564.

Franck, E., & Nüesch, S. (2011). The effect of wage dispersion on team outcome and the way team outcome is produced. *Applied Economics, 43(23)*, 3037-3049.

Franck, E., Verbeek, E., & Nuesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting, 26*, 448-459.

Frick, B., Prinz, J., & Winkelmann, K. (2003). Pay inequalities and team performance: Empirical evidence from the North American major leagues. *International Journal of Manpower, 24*, 472-488.

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under Social pressure. *The Review of Economics and Statistics, 87*, 208-216.

Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting, 26*, 460–470.

Jewell, R. T., & Molina, D. J. (2004). Productivity and salary distribution: The case of US major league baseball. *Scottish Journal of Political Economy, 51*, 127-142.

Katayama, H., & Nuch, H. (2011). A game-level analysis of salary dispersion and team performance in the national basketball association. *Applied Economics, 43*, 1193-1207.

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

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

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets? *The Economics Journal, 114(495)*, 223-246.

Livesport Media Ltd. (2019). *Odds Portal - Betting Odds Monitoring Service*. Retrieved from Oddsportal.com: https://www.oddsportal.com/

Martin, J. (1981). Relative deprivation: a theory of distributive injustice for an era of shrinking resources. *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews, 3*, 53-107.

Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting, 34*, 17-19.

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

Scully, G. W. (1974). Pay and Performance in Major League Baseball. *The American Economic Review, 64(6)*, 915-930.

Simmons, R., & Berri, D. J. (2011). Mixing the princes and the paupers: Pay and performance in the National Basketball Association. *Labour Economics, 18*, 381-388.

Sommers, P., & Quinton, N. (1982). Pay and Performance in Major League Baseball: The Case of the First Family of Free Agents. *The Journal of Human Resources, 17(3)*, 426-436.

Szymanski, S. (2015). *Money and Soccer.* New York: Nation Books.

Szymanski, S., & Smith, R. (1997). The English Football Industry: profit,performance and industrial structure. *International Review of Applied Economics, 11(1)*, 135-153.

Szymanski, S., & Zimbalist, A. S. (2002). Testing Causality Between Team Performance and Payroll. *Journal of Sports Economics, 3(2)*, 149-168.

Weissman, J. (2014). *Is Illegal Sports Betting a $400 Billion Industry?* Retrieved from http://www.inc.com/slate/jordan-weissmann-is-illegal-sports-betting-a-400-billionindustry.