Contract Year Effect in the NBA*

Changhao Shi† Jonathan Liu‡ Terry II Culpepper§ Sean Choi¶

June 12, 2020

Abstract

We investigate whether being in a contract year has an effect on NBA players by motivating them to put in more effort on playing. We use player statistics such as distance covered on the field per minute as a proxy for effort. We tackle three main challenges arising from the data: lack of a good proxy for effort, inter-correlation of effort between players, and player-specific effects on player statistics and outcomes. Using an OLS regression controlling for individual fixed effects and a double LASSO regression, we found that being in a contract year has no statistically significant impact on various measures of player effort.

Keywords: Sports Economics, Athletes, Contracts

JEL Codes: Z20, Z22

*We would like to thank Dr. Hortaçsu and Francisco del Villar Ortiz Mena for their unique insights.
†University of Chicago
‡University of Chicago
§University of Chicago
¶University of Chicago

1 Introduction

test

2 Literature Review

When watching the NBA, one finds that superstars such as Kobe Bryant, LeBron James, and Michael Jordan are not subject to the contract year phenomenon. This can be attributed to career concerns, as written by **?** and **?**. Such superstars play not simply to maximize their paychecks, but also to win championships, improve their legacies, and enter the conversation for greatest player of all time. However, many other players do not face the same career concerns, and hence may simply seek to maximize paychecks for minimal effort, forming the basis of the contract year phenomenon. At its core, the contract year phenomenon is a principal-agent problem, complicated by incomplete information and time-variant incentives. In other words, a principal pays an agent, where the principal reaps utility as a function of the effort the agent puts in, while the agent reaps utility from the payment. The cost to the principal is the payment, while that to the agent is his/her effort level. Furthermore, principals are unable to observe the effort levels of agents directly, meaning that contracts offered by the principal may be incentive-incompatible. This can then lead to what **?** call ”shirking” behavior, where an agent works with less effort than agreed upon in the contract. **?** further addresses the issue of moral hazard in a principal-agent interaction with imperfect information, and states that ”[w]hen the same situation repeats itself over time, the effects of uncertainty tend to be reduced and dysfunctional behavior is more accurately revealed, thus alleviating the problem of moral hazard”. However, a major issue with the contract year phenomenon is that principals, holding the expectation of high effort levels, offer long-term contracts, hence not allowing the ”same situation” to play out numerous times, but only a few times before the player retires. This is intuitively suboptimal, and in fact, **?** finds that ”contracts predicted by the theory” have not been well-confirmed empirically.

Thus, on the strategic level, agents on a contract year can engage in ex-ante opportunism, before undertaking ex post opportunism after signing the contract, terminology discussed in **?**. Namely, agents can first exert a high level of effort in one's contract year - ex ante opportunism - which enables them to sign a long-term contract. However, after signing, the players can now act with ex post opportunism, i.e. in this case putting in minimal effort. The degree to which this effort is minimal depends on an agent's dynamic problem. In other words, they can choose to exert no effort, hence guaranteeing that they never sign a new contract again. Alternatively, they could exert some positive level of effort, such that principals will be willing to give another contract once the agent reaches his next contract year. This is within the scope of the large body of IO literature on reputation.

That being said, sports-specific research has yielded conflicting results on opportunistic behavior.

3 Data

The NBA salaries of every player from the 2016-2017 season to the 2019-2020 season were collected from basketballreference.com. Note that basketballreference only displays data for the current season, so to get salary data for 2016-2017 to the 2018-2019 seasons, we utilized web.archive.org. Players were determined to be contract years if they did not have a salary for the following season. Furthermore, panel data of all advanced stats for each active NBA player was scraped from basketballreference.com for every regular season these players played in. Most of the advanced stats are measurements of a player's productivity on the court. These include the age of the player, team the player played for, position the player played for, the number of games the player played for a particular season, average minutes played per game, player efficiency rating, true shooting percentage, three-point attempt rate, free-throw attempt rate, offensive rebound percentage, defensive rebound percentage, total rebound percentage, assist percentage, free throw attempt rate, offensive rebound percentage, defensive rebound percentage, assist percentage, steal percentage, block percentage, turnover percentage, usage percentage, offensive win shares, defensive win shares, win shares, win shares per 48 minutes, offensive box plus/minus, defensive box plus/minus, box plus/minus, and value over replacement player.

Another way to measure a player's productivity/effort on the court is with boxout data. In theory, the more a player boxes out opposing players in order to grab rebounds, the more productive that player is to the team. All boxout-related data was collected from stats.nba.com for every player from the 2016-2017 to the 2019-2020 regular season. These include the number of boxouts a player averages per game, number of boxouts on offense a player averages per game, number of boxouts on defense a player averages per game, average number of rebounds the team grabs as a result of a player boxing out, average number of rebounds the player grabs as a result of him boxing out, percentage of times a player boxes out on offensive, percentage of times a player boxes out on defense, the percentage of times the teams grabs a rebound when boxing out, and the percentage of times the player grabs a rebound when boxing out.

Another method of measuring a player's value on the court is by using data on how often the player touches the ball. Players who handle the ball more often are typically much more valuable to his team. Data for player touches was collected from stats.nba.com. These include the number of touches a player averages per game, the number of touches the player averages in the front court per game, the percentage of the time the player has the ball when he is on the court, average seconds the player has the ball when he touches it, the average number of dribbles the player takes whenever he touches the ball, the number of points the player scores per touch, the average number of times a player touches the ball in the elbow part of the court, the average number of post-ups a player has per game, the average number of times a player touches the ball in the paint, points per elbow touch, points per post-up, and points per paint touch.

Note that when analyzing the data, the datasets for player salary, advanced player stats, boxouts, and touches were merged together by player name.

4 Model and Analysis

4.1 Heteroskedasticity

In general, we cannot expect that the data for NBA players follows a homoskedastic trend. For example, the variance for win shares for players on a contract year can theoretically be different than the variance for offensive win shares for players not on a contract year. This is because a contract year can cause behavioral shifts that are reified in different ways: while a contract year could push one player to prioritize their win shares to impress future teams, another may feel ready to retire and gradually decrease the effort they put into each match, therefore affecting their performance. In a similar vein, we would also expect that, all else equal, players who have more time on the field to have a lower variance in performance than players who have less time on the field. An argument on the Law of Large Numbers applies here: all else equal (as in, players have the same ability and are given the same opportunities), players' performance converge to their expected performance as t, the time they spend on the field, goes to infinity. Heteroskedasticity in this sense may cause the OLS estimator to be unbiased but inconsistent, possibly leading to invalid inferences based on biased variances, so verifying whether the data contains underlying heteroskedasticity and tackling the problem will help with lowering the variance and therefore the precision of our results.

4.1.1 Breusch-Pagan test

We begin by looking at whether the data shows heteroskedastic trends. To do so, we use the Breusch-Pagan test from Breusch and Pagan (1979). Our null hypothesis is

$$H_0 : \quad x_i \text{ exhibits homoskedastic trends}$$

for each regressor x_i we are using. To do so, we assume that the variance follows a functional form $h(z_i'\alpha)$. Then, our null hypothesis would be $\alpha = 0$, where 0 is the zero vector. Breusch and Pagan showed that this is equivalent to using the Lagrangian multiplier test statistic LM, where

$$\mathrm{LM} = \hat{d}'\hat{\mathcal{P}}^{-1}\hat{d}$$

and

$$d = \frac{\partial l}{\partial \alpha}$$

and

$$\mathcal{P} = -\mathbb{E}\left(\frac{\partial^2 l}{\partial\alpha\partial\alpha'}\right)$$

and the hats are quantities that are evaluated with $\hat{\alpha}$ (Breusch and Pagan (1979)). In **R**, we use the Bresuch-Pagan test on our OLS regression of win shares on contract year and our controls, and obtain a p-value of 6.564×10^{-5}:

```
hetero.plot <- lm(formula = ws ~ contract_year + as.factor(name) + pos +
    as.factor(season) +salary_current, data = nba)


bptest(hetero.plot)
```

Table 1: Studentized Breusch-Pagan test

	Test Results		
Regressand	BP	df	p-value
Win Shares	501.44	386	6.564×10^{-5}
Usage Rate	580.35	386	5.266×10^{-10}
Boxouts	530.19	386	1.433×10^{-6}
Average Speed	548.27	386	9.901×10^{-8}
Average Dribbles per Touch	664.3	386	0

The Breusch-Pagan test allows us to reject the null hypothesis that the dataset exhibits homoskedastic trends very confidently. Due to this result, we would require a method to tackle the underlying heteroskedasticity of our data.

4.1.2 Weighted OLS

A weighted OLS gives us a partial solution to this issue. Because we have reason to believe that observations of players who are given little playtime are of a worse quality than observations of players who are given plenty of playtime, a possible weighting mechanism is to use the minutes they spend per match in order to prioritize the observations of players who spend the most amount of time on the field. There is, of course, an issue of whether players who are given little playtime are different from players who are given plenty of playtime. We will turn to tackle this issue in this next section.

We first note that, *if* our weights are are inversely proportional to the underlying variance of the data (in other words, W, our weight matrix, is the inverse of the variance-covariance matrix), *then* the weighted OLS estimator is the BLUE estimator (Aitken (1936)). However, there is no reason to believe that the variance *is* the number of minutes in our data. Instead, it is reasonable to suppose that the variance *correlates* with the number of minutes. The potential disadvantage of a weighted least squares method arises when the weights are not precise relative to one another (Carroll and Ruppert (1988)), but because minutes played is an external measurement that is precise and are not estimated from within our data, we can be confident that the issue of weights being imprecise relative to one another is limited. To formalize our argument, we state our assumption.

Assumption 1 *W, where W is the weight matrix consisting of the minutes played for a given player for each observation, approximates the inverse of the variance-covariance matrix of the regression.*

Our regression model is therefore

$$y_{it} = x_{it}'\beta + \epsilon_i$$

but we estimate β with the weighted least square estimator

$$\hat{\beta} = \left(X'WX\right)^{-1} X'WY$$

where X is the data matrix, Y is the vector of y_{it}'s, and W is the weight matrix. To be clear, W is the matrix that consists of the minutes played for a given player for each observation, such that $X'WX$ outputs the data matrix $\hat{X}'\hat{X}$ where $\hat{X}$ consists of, for each observation i and time period t, $w_{it}x_{it}$ This shows that we need to make the following assumption:

Assumption 2 *$X'WX$ is an invertible matrix.*

We would have to argue that we do not run into the problem of perfect multi-collinearity. This is a reasonable assumption - we do not believe, for example, that when we weight the players' age and income by the minutes played per game on the field, we obtain a linear relationship between the two. Concerns would arise if we have covariates such as total distance run and speed together. Note that these two covariates may not initially have a linear dependence, until, possibly, when we weight them by minutes played. However, we haven taken care to prevent possible multicollinearity by avoiding using covariates that have some relationship with minutes played, or may relate to another variable after being weighted by minutes played.

4.2 Inter-correlation of observations

4.3 Player fixed effects

4.4 Team fixed effects

4.5 Curse of dimensionality and double LASSO

Challenges in estimating the contract year effect fall under three categories. Firstly, player effort cannot be observed, and existing metrics such as distance covered on the field per minute may only be loosely correlated with player effort. For example, a player may put in effort by perfecting 3-point throws and so roughly run the same amount of distance but attempt more 3-point shots per game. Another player may put in effort by running longer each game to gain tactical advantage on a field. It is also possible that the player's effort fails to translate into an observable metric; for example, he might put in effort into team-building exercises and coordinate much more on the field. This also brings us to the next challenge. The effort put in by a particular player may also correlate with other players' performance. Since basketball is a team game, effort put in by one player may synergize with effort put in by other players. We can also look at this issue game-theoretically. Consider the standard public goods game, and let y_i be the energy of the player i devoted to the match, out of a total of 1 endowed unit of energy. The payoff function for player i can be thought of as

$$\Pi_i = (1 - y_i) + \alpha \sum_j y_j$$

where $0 < \alpha < 1$. The sum of the players' effort correlates positively with the expected probability of winning a match. When $\alpha < 1$, the Nash equilibrium of this game is for all players to invest 0 energy into the match. However, when factoring in social norms, we expect players to fall into two categories. Based on empirical evidence, players will either put in more of their endowment when others put in more, or put in less when others put in more (?). Regardless, this makes estimating the contract year effect more difficult, as the effort devoted by an individual correlates with the effort of other individuals. Furthermore, this correlation is ex-ante unknown. The final challenge lies in the heterogeneity in different individuals. Even if we assume that the first two challenges are resolved, and that all players reflect their effort by pursuing, for instance, more three-point shots per minute, underlying unobserved parameters such as player skill and talent, player psychology during matches, and the correlation of the contract year effect with other covariates such as current salary means that individual fixed effects cannot be ignored.

To counteract the first challenge, we run multiple regressions with different outcome variables, such as player distance traveled per minute on the field and three point shot percentage. This partially circumvents the issue that proxies could be loosely correlated with player effort by looking at how being on a contract year affects a wide range of player statistics, instead of an individual measurement. To counteract the second challenge, we have to assume that the underlying correlation between players due to effort is homogeneous across teams. Then we can estimate the contract year effect on teams by looking at how many members are in a contract year and look at team level statistics. Finally, to counteract the third challenge, we control for fixed effects in our regression by adding indicator variables for players or teams as a covariate.

4.6 Controlling for Fixed Effects

To control for fixed effects (omitted variables of an individual), we add indicator variables for players or teams as a covariate. This, along with the existing controls, gives us the regression model

$$y_{it} = \sum_{j=1}^{N} \alpha_j \mathbb{1}\left\{i = j\right\} + \mathbb{1}\left\{i_t \in \text{Contract}\right\}\beta + c_{it}'\gamma + u_{it}$$

where y_{it} is the outcome player statistics for player i in time period t, $\mathbb{1}\{i = j\}$ is the indicator variable for players, $\mathbb{1}\{i_t \in \text{Contract}\}$ is the indicator variable for whether a player is in a contract year in a given time period, and c_{it} are various controls, such as age and current salary. The identifying assumption is then that unobservable effects soaked up by the individual indicator variable that simultaneously affect the outcome player statistics and the explanatory variable and covariates are time-invariant. In other words,

$$\text{Cov}\left(x_{i1}, u_{it}\right) = \cdots = \text{Cov}\left(x_{iT}, u_{it}\right) = 0,$$

where x includes both the explanatory variable and the controls. This assumption is innocuous enough in this context. We don't expect unobserved characteristics (omitted variables) of an individual to change dramatically across years that also affect whether a given player is in a contract year. For example, a player may get married and be very happy that year. This would cause him to play better, therefore increasing his performance and we will see a change in his player statistics. However it is unlikely that it will affect whether he is in a given contract year, as that aspect largely depends on how many years ago the player signed the contract.

5 Results

5.1 Player fixed effects

5.2 Team fixed effects

5.3 Double LASSO

6 Discussions

7 Conclusion

References

A. C. Aitken. IV.—On Least Squares and Linear Combination of Observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1936. ISSN 0370-1646. doi: 10.1017/S0370164600014346. URL https://www.cambridge.org/core/journals/proceedings-of-the-royal-society-of-edinburgh/article/ivon-least-squares-and-linear-combination-of-observations/7106C26F19F2EBF75BCEE7FA285780B9.

T. S. Breusch and A. R. Pagan. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5):1287–1294, 1979. ISSN 0012-9682. doi: 10.2307/1911963. URL https://www.jstor.org/stable/1911963.

R. J. Carroll and D. Ruppert. *Transformation and Weighting in Regression*. Springer US, Boston, MA, 1988. ISBN 9780412014215 9781489928733. doi: 10.1007/978-1-4899-2873-3. URL http://link.springer.com/10.1007/978-1-4899-2873-3.

Tables

Figures

Appendix A. Placeholder