Play for Pay: Quantifying the NBA's Contract Year Phenomenon Through Regression Analysis

Arnav Gurudatt
University of California, Berkeley

Stat 230A Final Project

10 May 2024

Video Link: https://tinyurl.com/yb9f24wx

Abstract

The existence of a "contract-year" phenomenon—where sports players are said to boost their performance to secure lucrative contracts in the following season—is widely debated across sports media. This phenomenon is said to be followed by a significant decline in performance once players secure such contracts. In this report, I analyze the NBA's encounter with this phenomenon, leveraging Python's webscraping libraries to create a novel dataset which combines advanced player performance metrics from FiveThirtyEight's Robust Algorithm (using) Player Tracking (and) On/Off Ratings (RAPTOR) index, historical salary data from HoopsHype, transaction data from RealGM, and comprehensive player statistics from Basketball Reference. I use Ordinary Least Squares (OLS) regression to estimate the association between contract years and player performance. Three key findings emerge: first, after controlling for confounders, players in contract years modestly outperform their non-contract-year counterparts by 0.25 to 0.3 points in the RAPTOR index; there is a statistically significant difference in offensive performance metrics, though no significant difference is found in defensive performance. Second, there is negligible change in performance the season after a contract year compared to non-contract-year players. Lastly, to address potential endogeneity issues highlighted in prior research, I use instrumental variables and two-stage least squares estimation, affirming a causal relationship between contract years and performance enhancements: since 2016, NBA players are expected to increase their total RAPTOR by 6.97 points, or 1.3 standard deviations, in their contract year.

Keywords: Contract year phenomenon, sports analytics, regression analysis, instrumental variables

1 Introduction

There is much discourse among sports talking-heads about the so-called "contract year" phenomenon – the idea that when a sports player knows they are due for a large new contract in the following season, they ramp up their production, hoping to boost their numbers in an effort to command more money the following year. The corollary to this phenomenon is that supposedly, after being signed to hefty contracts following a single-season surge in performance, these "contract-year" players notoriously see a drop in their production, to the chagrin of the team that signs them to maximum money. The National Basketball Association (NBA) is no stranger to this phenomenon, with plenty of notable examples in recent years. Former NBA center Erick Dampier is seen as the canonical example in sports media exemplifying the contract year phenomenon. After averaging career highs of 12.3 points and 12.0 rebounds per game in the final year of his contract with the Golden State Warriors in 2004, many teams were eager to sign him to their team the following summer. After signing a new contract with the Dallas Mavericks, however, Dampier would only average 9.2 points and 8.5 rebounds, before stringing together a couple of lackluster seasons averaging under 6 points per game with 4 rebounds and retiring after the 2012 season.

Despite the phenomenon being so widely known and discussed in sports media, there is very little existing literature actually quantifying it. Jean (2010) used a fixed-effects regression and found very large increases in points per game for players in their contract year; however, Jean himself concedes that his model has very "weak predictive power" with a negative adjusted R-squared value. White and Sheldon (2013) uses data from the 2003 through the 2012 seasons and finds that tracking player performance in 3-year windows (pre-contract year, contract year, post-contract year), there is a 0.5 points per game boost in the contract year followed by a 0.7 points per game decline the following year. Ryan (2015) was the first paper to estimate the true causal effect of the contract year on player performance, finding a “3-5 percentile boost in performance for the median player in the NBA” in the contract year.

However, there are several flaws in the existing literature, many of which are discussed at length in Ryan (2015). The first flaw is deciding on an accurate measure of player performance. Many papers try to use a single, aggregated statistical index of player performance, such as Player Efficiency Rating (PER) in White and Sheldon (2013). Considering PER alone is a very flawed approach, as this metric inflates the importance of big men in the NBA. Because PER weights rebounding so heavily, power forwards and centers, the tallest players in the NBA, have an undue advantage in this statistic which doesn’t accurately capture their performance and value on the court. Additionally, PER really only is a measure of efficiency on the offensive (scoring) side of the ball, with little consideration for defense. As a whole, no single, catch-all index will accurately capture all the intricacies of a very complex sport, but the existing literature does not analyze the best metrics and does not consider the robustness of their findings to alternative player performance indices.

The second problem brought up in Ryan (2015) is there is an induced bias in the way existing papers calculate the “contract year effect.” They analyze the change in performance for players who are on the last year of their contract and then sign a new contract the following season; but many players don’t actually go on to sign a new contract following a contract year and end up leaving the NBA. Because of this, the estimates of the effect of the “contract year” are upwardly biased. Players who played relatively poorly in their final contract year and do not get re-signed are not considered, effectively censored from the dataset.

The third problem is slightly more subtle. Ryan (2015) points out that there is an inherent endogeneity problem to regressing player performance on contract year status: better, star players sign longer-term deals, and are therefore less likely to be in a contract year compared to less impressive role players whose contracts expire every few years, which produces biased estimates of the contract year effect. While previous papers ignored this endogeneity, Ryan (2015) made use of instrumental variable regression to ascertain an actual causal effect of being in a contract year on player performance.

My work aims to expand upon the existing body of knowledge by constructing a comprehensive analysis that addresses these identified gaps. Rather than use one single metric for player performance, I focus on one robust advanced metric (RAPTOR) but also identify differences in contract year performance across a variety of offensive and defensive metrics. To reduce bias of omitting players who "drop-out" out of the league, I augment my dataset using a machine learning model to predict hypothetical performance metrics for players who otherwise left the league. Finally, while the primary focus of my paper is not causal inference, I attempt to replicate the instrumental variable regression from Ryan (2015) and provide a causal estimate of the effect of being in a contract year on an individual player's performance.

2 Dataset Construction

There is no one unified dataset containing statistics on player performance as well as information about players' contracts. In fact, defining a "contract year" is itself prone to ambiguity in the literature. Jean (2010) and White and Sheldon (2013) define a contract year as the year before a player reaches "free agency," or the ability to go sign a new contract with any team. While that definition reasonably encapsulates free agency, it omits a lot of other nuance to the NBA contract life cycle as pointed out in Ryan (2015). It completely omits players who do not sign a contract following the last year of their contract; it does not distinguish between "restricted" and "unrestricted" free agency; and it does not take into account the fact that players can sign contract extensions while their current contract is still active – which many star players elect to do, which is how they sign long-term deals without entering free agency. Even Ryan (2015)'s analysis ignores the fact that anticipating a contract extension can be reasonably considered a contract year; players have prior knowledge of when they are eligible to sign contract extensions for more money, and knowledge of that eligibility influences performance much in the same way impending free agency would. Thus, I define a contract year as a year in which (1) a player is eligible to sign a new contract after the year ends, or (2) a player signs a contract extension after the year ends while still on their current, non-expired contract.

I constructed a custom dataset leveraging data from four different sources that I webscraped using Python libraries. First, I was able to retrieve so-called "advanced analytics" player metrics from the sports analytics publication FiveThirtyEight. This dataset contained my primary player performance metric of interest: the Robust Algorithm (using) Player Tracking (and) On/Off Ratings (RAPTOR) index. RAPTOR measures the number of points a player contributes to team offense and team defense per 100 possessions, relative to a league-average player, and is considered to be one of the consensus "best metrics" among actual team managers and front offices to evaluate player performance. In addition to RAPTOR, this dataset also contains the Wins Above Replacement (WAR) metric from Ryan (2015). This measures how many wins a player contributes to their team relative to a "replacement-level player" that the team could add with very little effort. These metrics were indexed by player and by season. The schema for this dataset is below.

Column Name	**Type**	**Description**
player_id	String	Unique identifier for each NBA player.
player_name	String	Name of the NBA player.
season	Integer	NBA season year.
poss	Integer	Total possessions played by the player during the season.
mp	Integer	Minutes played by the player during the season.
raptor_offense	Float	RAPTOR offensive rating per 100 possessions.
raptor_defense	Float	RAPTOR defensive rating per 100 possessions.
raptor_total	Float	Total RAPTOR rating, combining offense and defense.
war_total	Float	Wins Above Replacement total for the player.

Table 1: Data Schema for NBA Player Performance Metrics

The next data source was salary data from the website HoopsHype. Contract salaries are available starting from the 1990-91 season up until the current season, and come both as nominal salaries as well as inflation-adjusted salaries in current US dollars. I used HTML parsing from the Python library pandas to iterate through each year's salary data. The schema for the webscraped and aggregated dataset from each year is below.

Column Name	**Type**	**Description**
Player	String	Name of the NBA player.
Salary	Currency	Contracted salary for the season, not adjusted for inflation.
Real Salary	Currency	Salary adjusted for inflation to current year's dollars.
Year	Integer	The NBA season year for which the salary data applies.

Table 2: Data Schema for NBA Player Salary Information

The third data source was transactions data from the website RealGM. I used the text in the transaction data to determine whether a year was a contract year or not. My definition of a contract year can essentially be boiled down to when a player anticipates the opportunity for a new (often larger) contract the following season. As such, if a player signed a contract in one season, it was reasonable to infer that the previous season was a contract year. RealGM uniformly tracks whenever players sign contracts with the following text: "Date - Player Name signed a contract with the Team Name" if the signing happened in free agency; "Date - Player Name signed a multi-year contract with the Team Name" for mid-season contract extensions; "Date - Player Name signed a rookie scale extension with the Team Name" for rookie contract extensions; and "Date - Player Name signed a veteran extension with the Team Name" for veteran extensions. This scraping task essentially boiled down to using HTML parsing to retrieve player data and the transactions table, and then using string matching to get the year of the transaction find instances of the word "sign"; if a player "signed" any sort of contract at a given point in time, then the previous season was taken as a contract year (if a player signs a new contract in 2019, then the 2018-19 season is a contract year). Of course, this method is not immune to the same bias described in Ryan (2015)'s paper: it omits players who did not get a contract following their last contract year. However, I was able to rectify this issue using data imputation and augmentation with machine learning, which I will describe in the next section. Scraping this website for player transaction data took around 12 hours, and I was only able to retrieve information for about 80% of the players in the HoopsHype and RAPTOR datasets. After retrieving all the information from the HTML pages, I applied some data preprocessing techniques to extract the player name, year, and position from the data, and created a binary indicator column for whether that player was in a contract year or not based on whether they "signed" a contract the following season. The schema for the transactions table is below.

Column Name	**Type**	**Description**
Transaction	String	Description of the transaction event including dates.
Player	String	Name of the NBA player involved in the transaction.
Position	String	The position played by the NBA player.
Year	Integer	The NBA season year related to the transaction event.
Contract Year	Integer	Binary indicator (1 or 0) where 1 indicates the transaction comes after a contract year.

Table 3: Data Schema for NBA Player Transaction Events

Lastly, I was interested in measures of player performance besides those in FiveThirtyEight's RAPTOR dataset. Basketball Reference is an online database that contains countless player and team performance statistics dating back to the very first NBA season in 1947. I downloaded player per-game statistics, statistics adjusted per 100 possessions[1], and advanced statistics such as Win Shares (WS), Box Plus-Minus (BPM), and Value Over Replacement Player (VORP) from 1991

[1] In basketball, a "possession" refers to the control of the ball by one team who then attempts to score. Statistics adjusted per possession offer a more standardized way to compare players, accounting for the pace of different games and teams.

to 2022. The schema for this dataset is below.

Column Name	**Type**	**Description**
player_additional	String	Unique identifier for each player.
Player	String	Name of the NBA player.
Pos	String	Player's position on the team.
Age	Integer	Age of the player during the season.
Tm	String	NBA team the player was part of for the season.
Box Score Statistics	Numeric	Standard basketball statistics including games played (G), games started (GS), minutes played (MP), field goals made (FG), field goal attempts (FGA), 3-point field goals made (3P), 3-point attempts (3PA), 2-point field goals made (2P), 2-point attempts (2PA), free throws made (FT), free throw attempts (FTA), offensive rebounds (ORB), defensive rebounds (DRB), total rebounds (TRB), assists (AST), steals (STL), blocks (BLK), turnovers (TOV), personal fouls (PF), and points scored (PTS).
Advanced Statistics	Numeric	Advanced player performance metrics including offensive win shares (OWS), defensive win shares (DWS), total win shares (WS), offensive box plus-minus (OBPM), defensive box plus-minus (DBPM), box plus-minus (BPM), and value over replacement player (VORP).
Season	Integer	NBA season year.

Table 4: Data Schema for NBA Player Statistics

I constructed one final, merged dataset by joining the tables on matching player names or player IDs as well as season. The final dataset is similar to an unbalanced panel dataset, where we have observations for individuals (NBA players) over time, but we do not have observations for the same individuals in each year of the dataset, due to naturally arising factors like entering the NBA or player retirement.

3 Addressing Bias: A Machine Learning Application

As previously discussed, Ryan (2015) highlights the biased induced by omitting players who do not sign a new contract after their final contract year. These players have anticipatory knowledge of being able to sign a new contract after the year concludes, yet, due to potentially bad performance or retiring in old age, these players do not sign a contract the following season. This is potentially a source of bias in the estimates for player performance post-contract year: if players performed so poorly in their contract year that they do not even get signed to a new contract, effectively "dropping out" of the NBA, then we are only measuring the change in performance of players who stayed in the league – who are more likely to be "good" players. I come up with a solution to this problem using two different methods. While it is impossible to know definitively how a player would have performed in a hypothetical post-contract year, we can estimate their performance under some reasonable assumptions.

3.1 Data Imputation

I first opt for a very naive imputation approach. I find the final season played by each NBA player, and then add an additional row for each player, incrementing the player's age and season by one accordingly. This sets the groundwork for creating a hypothetical final season played after the player's true final season. I use two imputation approaches: in the first one, I impute the absolute minimum amount of production in the hypothetical final season. For counting stats that are necessarily non-negative such as points per game or minutes played, this would be an imputed value of 0. For advanced statistics that can take either positive or negative values like RAPTOR or BPM, I imputed the player's career low in that category. Effectively this minimum imputation approach can be boiled down to $\min(0, \text{Career Low}_{iC})$ for player i in statistical category C. This imputation method is not meant to be a realistic estimate of player performance; it is meant to be an absolute lower bound estimate of a player's performance that provides estimates assuming an absolute worst case scenario for that player's performance in a hypothetical final season.

The second imputation approach was adding an additional row for each player that was identical to their actual final season in every statistical category (except for age and year, which were incremented by 1). This is also a naive imputation approach, but it allows us to get estimates for how a player would have performed assuming there is no change in that player's performance from their actual final season. This is a fairly strong assumption to make, since players towards the end of their careers generally continue to decline in performance. However, these declines are usually not absurdly large in magnitude, and by leaving their performance identical to their actual final season, we are essentially estimating a case where their performance "gets no worse."

3.2 Data Augmentation with Machine Learning

The third approach I used to estimate a hypothetical "final season" was to predict how a player would have performed in their last season using machine learning models. The rationale for this approach is that while the imputed values would be synthetic, they would be realistic predictions of a player's performance. I used an auto-regressive linear model with a lag order of 3 to predict a given box score or advanced statistic based solely on values of that statistic (and other statistics) from previous seasons:

$$\hat{y}_t = \hat{\phi}_0 + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} + \hat{\phi}_3 y_{t-3} + \hat{\gamma}^T \Theta = X\hat{\Phi}$$

In this model specification, we are predicting a statistic $\hat{y}_t$ using three lagged values y_{t-1}, y_{t-2}, and y_{t-3}; Θ is a matrix of lagged versions of other box score and advanced statistics, as well as lagged salary, year, age squared, and dummies for position as features. (Equivalently, the regression model can be written as the matrix-vector product of the design matrix X and weight vector $\hat{\Phi}$.) I split the data sample into a training set on which I fit the model to "learn" the optimal weights, and held out 20% of the data as the test set on which I validated my model's ability to generalize to unseen data. All features except categorical features were standardized during the training and testing process, but inverted back to their original units after.

To reduce potential overfitting due to the high dimensionality of the feature-space, I trained three more models adding regularization terms to the base AR(3) model:

1. A LASSO model with L1 regularization ($\alpha = 0.01$), with $\hat{\Phi}$ given by:
   $\min_{\hat{\Phi}} \left(\frac{1}{2n} \|y - X\hat{\Phi}\|^2 + \alpha \|\hat{\Phi}\|_1 \right)$
2. A Ridge model with L2 regularization ($\alpha = 1$), with $\hat{\Phi}$ given by:
   $\min_{\hat{\Phi}} \left(\frac{1}{2n} \|y - X\hat{\Phi}\|^2 + \alpha \|\hat{\Phi}\|_2^2 \right)$
3. An Elastic Net model with both L1 and L2 regularization ($\alpha = 0.01$, L1 Ratio $\lambda = 0.5$), with $\hat{\Phi}$ given by:
   $\min_{\hat{\Phi}} \left(\frac{1}{2n} \|y - X\hat{\Phi}\|^2 + \alpha\lambda \|\hat{\Phi}\|_1 + \frac{1}{2}\alpha(1 - \lambda) \|\hat{\Phi}\|_2^2 \right)$

I elected not to train models that introduce non-linearity such as Random Forests or neural networks as they are out of scope for this course. I compared the root mean squared error (RMSE) and mean absolute error (MAE) of each model on predicting each statistic of interest in the test set, and then also averaged the RMSEs and MAEs of the four models to choose the overall best model to make predictions for my synthetically generated final seasons. The Ridge model had the lowest test RMSE while the Elastic Net had the lowest test MAE; I opted to use the Elastic Net model to predict player performances in hypothetical final seasons since its performance to the Ridge model was largely comparable. Mean model performance metrics are displayed below, but full RMSE, MAE, and sample standard deviation values for each predicted statistic can be found in the tables appendix.

Ultimately, I constructed four datasets: the standard dataset without any imputed values, the two naively imputed datasets, and the augmented dataset with synthetically generated values from the Elastic Net model.

Table 5: Mean RMSE and MAE for each model

Model	RMSE	MAE
OLS	0.738278	0.510616
LASSO	0.740506	0.510269
Ridge	0.737220	0.509500
ElasticNet	0.737887	0.508305

4 Summary Statistics

I was interested in analyzing the association between being in a contract year and the performance metrics of players measured by the following statistical measures: total RAPTOR, points per game (PTS), offensive win shares (OWS), defensive win shares (DWS), total win shares (WS), offensive box plus-minus (OBPM), defensive box plus-minus (DBPM), box plus-minus (BPM), value over replacement player (VORP), offensive RAPTOR, defensive RAPTOR, and wins above replacement (WAR). Descriptive statistics for each of these measures (in the non-imputed dataset) is below.

Table 6: Summary Statistics for Various Metrics

Metric	Count	Mean	Std Dev	Min	25%	50%	75%	Max
raptor_total	12491	-1.39	4.10	-63.75	-3.40	-1.24	0.88	47.47
PTS	12491	8.59	6.05	0.00	3.90	7.10	12.00	36.10
OWS	12491	1.47	2.12	-3.30	0.00	0.80	2.30	14.90
DWS	12491	1.35	1.24	-1.00	0.40	1.00	2.00	9.10
WS	12491	2.82	3.05	-2.10	0.40	1.90	4.30	20.40
OBPM	12491	-1.30	3.39	-41.40	-3.00	-1.10	0.60	42.10
DBPM	12491	-0.13	1.52	-26.10	-0.90	-0.20	0.70	17.80
BPM	12491	-1.43	4.07	-52.40	-3.20	-1.20	0.70	41.50
VORP	12491	0.69	1.42	-2.60	-0.10	0.10	1.10	11.80
raptor_offense	12491	-1.11	3.17	-43.37	-2.70	-1.00	0.62	53.23
raptor_defense	12491	-0.28	2.30	-43.21	-1.37	-0.34	0.77	62.47
war_total	12491	2.10	3.59	-5.43	-0.10	0.72	3.30	28.76

Notably, while league-average points per game have been slowly increasing since 2015, most statistical categories have remained relatively time-stationary since 1991. Even among points per game, the recent upward trend in scoring seems to be more of a reversion back to the levels of scoring in the 1990s. There were also slight upward trends in league-average total RAPTOR, offensive RAPTOR defensive RAPTOR, BPM, and OBPM from 2015 to 2020, but these metrics have since declined and reverted back to pre-2015 levels.

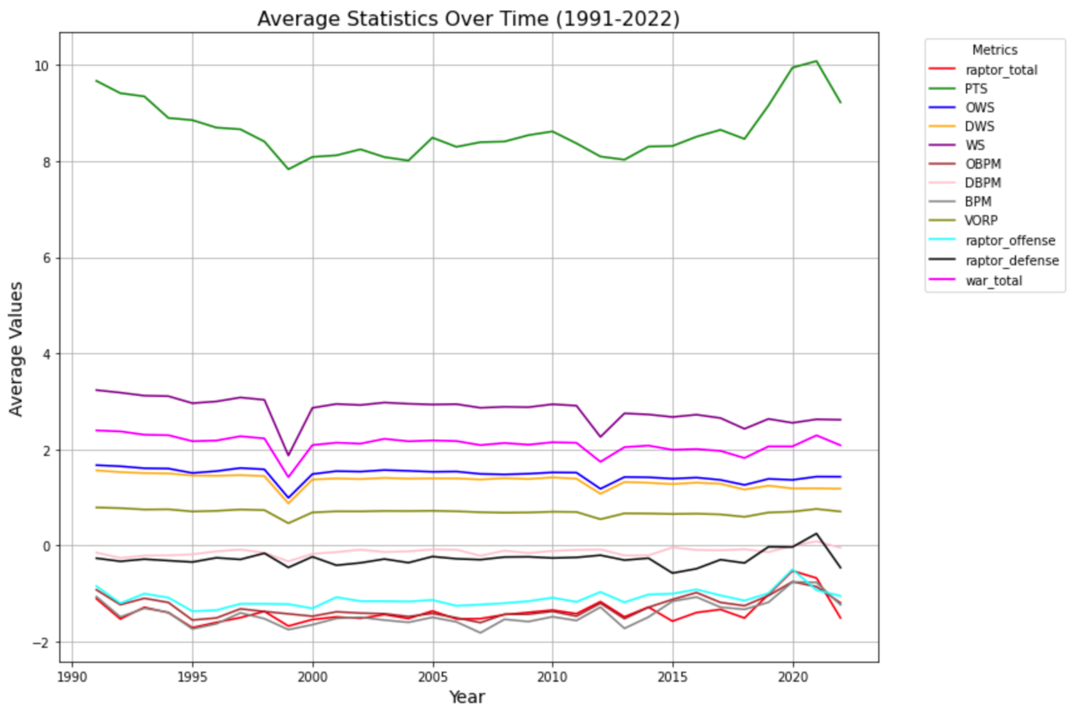

Figure 1: League-Average Value of Statistical Categories, 1991 to 2022

We can also examine how NBA player salaries have changed over time. After being stagnant for much of the 2000s following a period of growth from the 1990s, in 2016, the NBA Players' Association negotiated a new Collective Bargaining Agreement (CBA) that made unprecedented changes to NBA salary structure (Brown (2015)). It added additional sources of revenue that players would be entitled to a share of, leading to a projected $1.5 billion more in the first year of the new CBA compared to the previous CBA negotiated in 2011. Additionally, the NBA increased the salary cap – roughly the maximum amount teams are allowed to spend on all their players – from $70 million the previous year to $94.143 million in 2016-17, a 34.5% increase; by comparison, the NBA salary cap had only increased by 29.3% from 2007 to 2016, meaning that NBA team purses saw more growth in a single season than in nearly the previous decade. By 2022, the NBA salary cap had increased to $112.414 million.

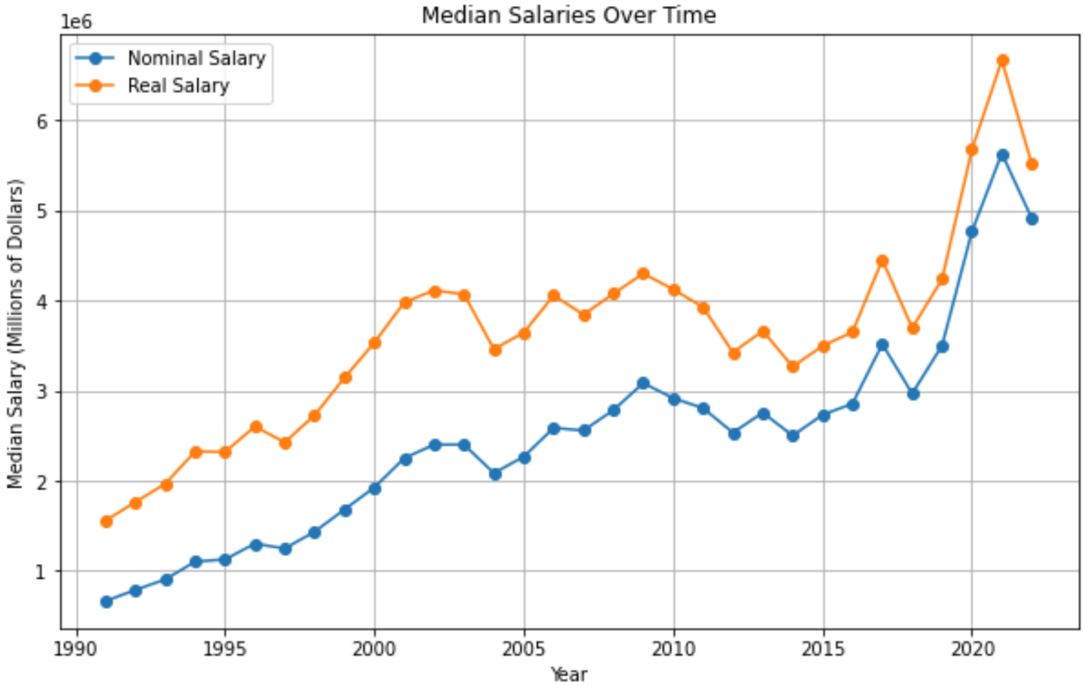

Figure 2: Nominal and Real NBA Salaries, 1991-2022

The following year, the 2017 CBA saw the introduction of the Designated Veteran Player Extension, known colloquially as the "Supermax" contract (Brazda (2023)); under this contract structure, star players were receiving contracts that reached sizes never seen before. According to the NBA year-by-year salary tracker Spotrac, in 2016, Kobe Bryant was the highest paid player in the league, earning \$25 million that season, and was the only player besides Michael Jordan to ever have a single-season salary of at least \$30 million; after the addition of the "Supermax" contract in the NBA, 10 players had a salary of at least \$30 million in the 2022 season. Golden State Warriors guard Stephen Curry became the first player to earn a single-season salary of at least \$40 million in 2020, and at least \$50 million in 2024. Boston Celtics player Jaylen Brown is projected to make more than \$65 million in 2029 after signing the largest contract ever in 2023. In short, NBA salaries have seen unprecedented levels of growth thanks to the 2016 CBA. Nominal salaries have grown by 642.06% from 1991 to 2022, and adjusted for current US dollars, they have grown 254.79% in the same time frame. Median salaries also grew 59% from 2019 to 2022 after minimal growth and even periods of decline in the previous two decades.

Table 7: Median Nominal and Real Salary Growth in the NBA, Four Year Windows

Period	**Nominal Salary**		**Real Salary**	
	Amount	Growth	Amount	Growth
1991-1994	845,000	—	1,882,448	—
1995-1998	1,297,500	53.55%	2,534,189	34.62%
1999-2002	2,000,428	54.18%	3,671,588	44.88%
2003-2006	2,392,500	19.60%	3,842,058	4.64%
2007-2010	2,765,900	15.61%	4,044,527	5.27%
2011-2014	2,626,473	-5.04%	3,499,518	-13.48%
2015-2018	2,996,390	14.08%	3,794,854	8.44%
2019-2022	4,767,000	59.09%	5,631,382	48.40%

Table 8: Summary Statistics for NBA Salaries in 1991 and 2022 (values in $millions)

Year	Count	Mean	Std Dev	Min	Median	Max
1991 Nominal Salary	338	0.843	0.678	0.030	0.663	4.250
2022 Nominal Salary	251	9.520	10.701	0.037	4.916	45.781
1991 Real Salary	338	1.982	1.567	0.070	1.556	9.983
2022 Real Salary	251	10.691	12.017	0.042	5.521	51.411

Finally, I examined share of NBA players in their contract year since the 1991 season. The large fluctuations in 1998 and 1999 are to be expected: the 1998-1999 NBA season saw a significant decline in players entering their contract year because of the NBA lockout, which delayed the start of the season and shortened it to just 50 games, compared to the usual 82. This lockout, rooted in contract and labor disputes, led to fewer contract renewals or negotiations during the year. Consequently, in 2000, there was a noticeable surge in contract year shares as many players' contracts that were due for renewal in 1999 were deferred to 2000, following the resolution of the lockout. This led to an unusually high number of players entering or renegotiating contracts at the same time, influencing the overall data trends seen in this period. Since 2011, the share of players entering their contract year every season has consistently been increasing. In past decades, it was common for players to sign much longer-term deals. Particularly, it used to be common for players to sign seven-year contracts with teams (Reina (2022)); Erick Dampier's infamous contract with the Mavericks was $73 million over seven years. However, the maximum contract length has since been limited to just four or five years for free agents due to CBA negotiations in 2011 and 2012. As a result, players are on teams for shorter periods of time, causing a greater share of them to enter contract years over time. In 1991, just under 29% of players were in their contract year, compared to 59% in 2021 and 47% in 2022.

Figure 3: Share of Players in their Contract Year, 1991-2022

Table 9: Average Share of NBA Players in Contract Year, Four-Year Windows

Period	**Share**
1991-1994	0.297610
1995-1998	0.261214
1999-2002	0.399624
2003-2006	0.346939
2007-2010	0.351090
2011-2014	0.424313
2015-2018	0.506795
2019-2022	0.533493

5 Modeling

5.1 Average Performance in a Contract Year

I first sought to quantify the change in expected performance that comes with being in a contract year. This essentially boils down to a difference in means comparison: if we regress player performance of player i in year t on a dummy variable indicating whether that player is in a contract year or not, i.e.:

$$\text{Performance}_{i,t} = \beta_0 + \beta_1 \text{Contract Year}_{i,t} + \gamma^T \mathbf{X} + \epsilon_{it}$$

where $\mathbf{X}$ is a matrix of covariates, then a positive β_1 would show that relative to players not in a contract year, players in a contract year have a greater expected performance; and conversely, a negative β_1 would show that players in a contract year have a worse expected performance compared to those not in a contract year.

I start with total RAPTOR as my measure of player performance and estimate the simple regression without covariates below, using White robust standard errors:

$$\text{raptor\_total}_{i,t} = \beta_0 + \beta_1 \text{Contract Year}_{i,t} + \epsilon_{it}$$

Seemingly contrary to the prevailing narratives around contract years, I find that the point estimate $\hat{\beta}_1$ is actually negative: on average, players who are in their contract year have a RAPTOR total that is 0.7368 points lower than players not in their contract year, an estimate which was statistically significant at the 0.001 level. I compare the difference in mean player performance across a variety of different measures, but arrive at the same finding each time. Player performance for players in their contract year is on average lower across every statistical measure to a statistically significant degree compared to non-contract year players.

There could be several reasons for this seemingly counterintuitive finding. The simplest explanation is that most NBA players are not star players, and as Ryan (2015) observed, non-star players are more likely to be in a contract year. Out of

Table 10: Difference in Mean Performance Across Various Dependent Variables

Dependent Variable	Estimate	Std. Error	P-value	95% CI
RAPTOR Total	-0.7368	0.078	<0.001	[-0.889, -0.585]
PTS	-2.3402	0.106	<0.001	[-2.548, -2.133]
OWS	-0.5749	0.037	<0.001	[-0.648, -0.502]
DWS	-0.4444	0.022	<0.001	[-0.487, -0.402]
WS	-1.0195	0.053	<0.001	[-1.124, -0.915]
OBPM	-0.7564	0.063	<0.001	[-0.881, -0.632]
DBPM	-0.0971	0.029	0.001	[-0.155, -0.040]
BPM	-0.8537	0.077	<0.001	[-1.005, -0.703]
VORP	-0.3412	0.025	<0.001	[-0.389, -0.293]
RAPTOR Offense	-0.5456	0.060	<0.001	[-0.663, -0.428]
RAPTOR Defense	-0.1912	0.045	<0.001	[-0.280, -0.103]
War Total	-0.9199	0.062	<0.001	[-1.042, -0.798]

Note: Standard errors are HC2 robust.

the 15 players on an NBA roster, only 8 to 10 are regularly playing in the rotation, and of those 8 to 10, perhaps 3 to 5 would be good enough to start on every other team in the league.

The R^2 for each of these models was considerably low, with no model having more than 4% of the variation in the performance metric explained by the regression on contract year. I introduce controls for minutes played and possessions to see how they affect the point estimate for change in performance in a contract year.

Table 11: Regression Results with Dependent Variables and Minutes Played (MP) as Control

Dependent Variable	Contract Year Coef. (Std. Err.)	MP	Adjusted R^2
RAPTOR Total	0.3081*** (0.065)	0.0023*** (0.00)	0.300
PTS	-0.1139 (0.066)	0.0049*** (0.00)	0.646
OWS	0.1620*** (0.025)	0.0016*** (0.00)	0.559
DWS	0.0176 (0.012)	0.0010*** (0.00)	0.661
WS	0.1789*** (0.029)	0.0026*** (0.00)	0.721
OBPM	0.1805*** (0.052)	0.0021*** (0.00)	0.356
DBPM	0.0256 (0.028)	0.0001*** (0.00)	0.031
BPM	0.2102*** (0.064)	0.0024*** (0.00)	0.319
VORP	0.1133*** (0.018)	0.0010*** (0.00)	0.476
RAPTOR Offense	0.2612*** (0.050)	0.0018*** (0.00)	0.299
RAPTOR Defense	0.0469 (0.043)	0.0005*** (0.00)	0.05
WAR Total	0.3085*** (0.043)	0.0027*** (0.00)	0.544

Note: *** p<0.001, ** p<0.01, * p<0.05. Standard errors are HC2 robust.

Table 12: Regression Results with Dependent Variables and Possessions (POSS) as Control

Dependent Variable	Contract Year Coef. (Std. Err.)	POSS	Adjusted R^2
RAPTOR Total	0.2906*** (0.065)	0.0012*** (0.00)	0.300
PTS	-0.1345* (0.065)	0.0025*** (0.00)	0.656
OWS	0.1477*** (0.025)	0.0008*** (0.00)	0.556
DWS	0.0055 (0.013)	0.0005*** (0.00)	0.649
WS	0.1526*** (0.029)	0.0013*** (0.00)	0.714
OBPM	0.1929** (0.064)	0.0012*** (0.00)	0.319
DBPM	0.0404 (0.043)	0.0003*** (0.00)	0.049
BPM	0.1929** (0.064)	0.0012*** (0.00)	0.319
VORP	0.1039*** (0.018)	0.0005*** (0.00)	0.472
RAPTOR Offense	0.2501*** (0.050)	0.0009*** (0.00)	0.301
RAPTOR Defense	0.0404 (0.043)	0.043*** (0.00)	0.049
WAR Total	0.2827*** (0.043)	0.0014*** (0.00)	0.540

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust.

With just one additional covariate, the coefficients on contract year for player performance as measured by total RAPTOR, OWS, WS, OBPM, BPM, VORP, offensive RAPTOR, and WAR are positive and statistically significant, suggesting that controlling for minutes played and possessions, players who are in their contract year actually perform better than their non-contract year counterparts. With total RAPTOR specifically, controlling for minutes played and possessions, players in their contract year saw increase of 0.29 to 0.3 in their RAPTOR index performance on average. However, the positive association between contract year status and player performance is interestingly only restricted to offensive and aggregate performance metrics; I observe no statistically significant difference in average player performance for contract year versus non-contract years across defensive metrics (DWS, DBPM, defensive RAPTOR), as well as no statistically significant difference in points per game between contract year and non-contract year players controlling for minutes played (but a small, statistically significant decrease at the $\alpha = 0.05$ level controlling for possessions). Taken together, these findings suggest that being in a contract year generally does have a positively associated change in player performance, and most of this positive association is driven by gains on the offensive end. However, it seems that even these offensive improvements do not really materialize in any observable "box score statistics" such as how many points per game a player scores. Additionally, being in a contract year seems to have no associated change in defensive performance. This makes sense intuitively; fundamentally, basketball is a game of scoring, and the players who score most tend to command the largest salaries. It would make sense that knowing they have the potential to make lots of money the following year, players pay special attention to their offensive production rather than defensive production.

To check for covariate sensitivity, I introduced further controls for position, age, standardized salary, and number of games started, and also compared this model to a model with additional year fixed-effects. Our primary performance metric of concern is total RAPTOR, so I include the full regression table for this metric with and without year fixed-effects (to control for potential time-related variation in the dependent variable), and relegate the other dependent variables' regression tables to the appendix.

Table 13: Regression Results for Total RAPTOR with Control Variables and Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-5.5939*** (0.259)	-5.6420*** (0.336)
Contract Year	0.2924*** (0.065)	0.2501*** (0.066)
Possessions (POSS)	0.0012*** (0.00003)	0.0012*** (0.00003)
Age	0.0195* (0.008)	0.0202* (0.008)
Games Started (GS)	-0.0082*** (0.002)	-0.0089*** (0.002)
Salary	0.5056*** (0.031)	0.4769*** (0.043)
Forward (F)	0.4774*** (0.101)	0.4641*** (0.101)
Guard (G)	0.9347*** (0.102)	0.9184*** (0.102)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.321	0.325
Adjusted R-squared	0.320	0.323
F-statistic	634.3	148.0
Prob (F-statistic)	0.00	0.00

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Here, we see that the statistically significant increase in mean total RAPTOR among contract year players is robust to covariate selection. Controlling for confounders, total RAPTOR is on average 0.2924 points higher among contract year players, or 0.2501 controlling for year-by-year variation with year fixed effects. With these additional covariates, I see similar, statistically significant positive increases in offensive and aggregated player performance metrics OWS (Table 24), WS (Table 26), OBPM (Table 27), BPM (Table 29), VORP (Table 30), offensive RAPTOR (Table 31), and WAR (Table 33), just as before, but no statistically significant difference in defensive metrics (Table 28) or points per game (Table 23) between contract year and non-contract year players.

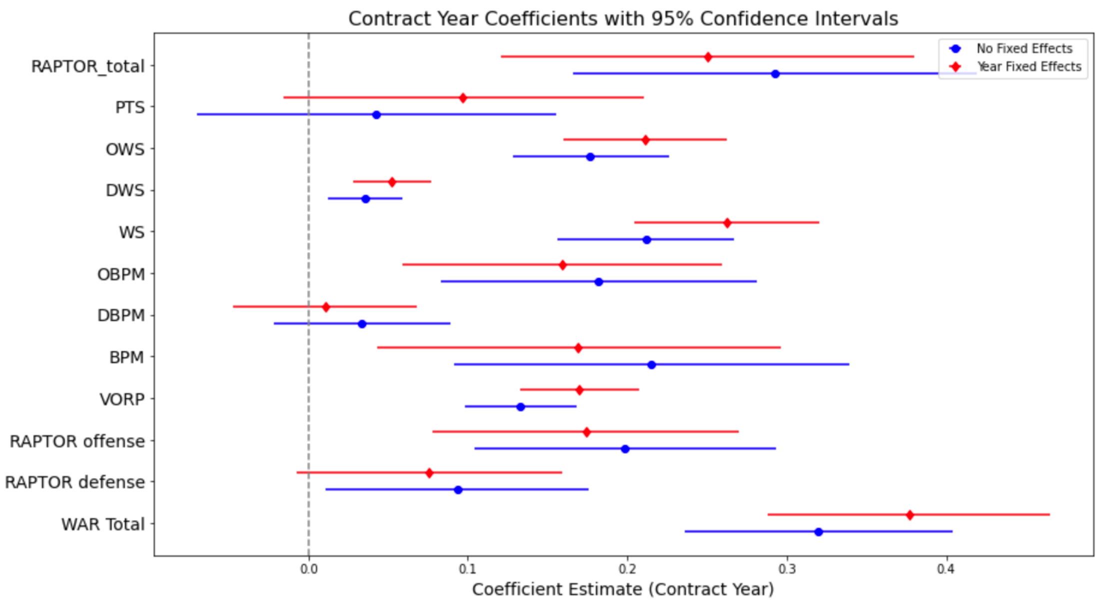

Figure 4: Point Estimates of Change in Performance for Contract Year Players

Thus, controlling for relevant confounders, being in a contract year is expected to positively increase player performance, particularly offensive performance, but there is no observed increase in defensive performance or points per game. Importantly, these findings are not causal, but rather a purely an association-based comparison of mean performance of players in contract years versus those not in contract years.

5.2 Change in Performance Post-Contract Year

I was then interested in knowing the "change" in performance of players just after their contract year relative to players who were not in a contract year in either time frame.

This is somewhat analogous to the "difference-in-differences" model. For example, let's say we want to know how player performance changed just after their contract year from the 2019 to the 2020 season, relative to players who were not in a contract year in either 2019 or 2020. If we treat a player being in a contract year as a quasi-treatment, while not being in a contract year is analogous to a control group, then if we estimated a difference-in-differences regression specification:

$$\text{Performance}_{i,2020} = \beta_0 + \beta_1 \text{Contract Year}_{i,2019} + \beta_2 \text{Y}_{2020} + \beta_3 \text{Contract Year}_{i,2019} \times \text{Y}_{2020} + \epsilon_{i,2020}$$

1. β_0 would represent the performance of players who were not in a contract year in 2019, in 2019

2. β_1 would represent the difference in performance of players who were in a contract year in 2019 relative to players who were not in a contract year in 2019, in 2019

3. $\beta_0 + \beta_2$ would represent the performance of players who were not in a contract year in 2019, in 2020

4. $\beta_0 + \beta_1 + \beta_2 + \beta_3$ would represent the performance of players who were in a contract year in 2019, in 2020. So $\beta_2 + \beta_3$ is the change in performance from 2019 to 2020 for players in their contract year in 2019

5. β_3 is the coefficient of the interaction term and is also our coefficient of interest - it represents the so-called **gain score**, or the change in performance in 2020 from 2019 for players who were in a contract year in 2019 relative to players who were not in a contract year in either 2019 or 2020.

It should be very important to note here that while the traditional difference-in-differences model crucially relies upon the parallel trends assumption to make causal claims about β_3, by definition, the parallel trends assumption is violated in our setup because players swap in and out of the "treatment" and "control" groups every year depending on whether they are due for a contract. Thus, it is impossible, and also incorrect, to say that the "treatment" and "control" groups had similarly trending performances pre-contract year, because in reality, no such treatment or control groups exist. As such, our estimate of β_3 is NOT causal evidence of any contract-year "effect". Rather, it merely represents the difference in performance change between players who just got out of their contract year versus players who were not in their contract year in 2019 or 2020 (from our example).

We can accordingly expand our model to take all years and interactions, along with covariate matrix $\mathbf{X}$, into account to estimate the year-by-year difference in performance change after coming off a contract year (CY). In this regression, we estimate a player i's performance y in year t coming off of a contract year in year $t-1$ and an indicator variable $I(t)$ which equals 1 if we are estimating the change in year t from year $t-1$ with t ranging from 1992 to 2022.

$$y_{i,t} = \beta_0 + \beta_1 \mathrm{CY}_{i,t-1} + \textstyle\sum_{t=1992}^{2022} \alpha_t I(t) + \sum_{t=1992}^{2022} [\delta_t \mathrm{CY}_{i,t-1} \times I(t)] + \gamma^T \mathbf{X} + \epsilon_{i,t}$$

In this section, I will also leverage my synthetically imputed datasets to produce less biased point estimates of change in performance.

My primary finding is that whether or not players "improve" or "regress" the year right after their contract year is largely a moot question, because whatever increase or decrease they experience in performance is largely no different than the change in performance of players who were not in a contract year in any pair of adjacent years. That is to say, even if it were the case that players uniformly declined in performance post-contract year, there would also need to be some explanation as to why non-contract year players also declined by approximately the same amount in the same time frame. Furthermore, the lack of change in performance relative to a "non-contract year" baseline still holds up even after correcting for the potential bias in the dataset where we omit players who did not sign a new contract after their final contract year. While this lack of a difference does not have a purely causal interpretation, it does lend credence to the idea that players' season-by-season change in performance seems to be unassociated with whether or not they are coming off of a contract year. Thus, the so-called "contract year phenomenon" seems to be based off of a couple of notable examples of players who declined after signing lofty contract extensions: Erick Dampier, Dion Waiters, and Andrew Wiggins come to mind. However, this ignores the fact that the majority of players generally maintain their level of performance without getting immediately better or worse following their contract year.

Comparing the estimates uncorrected for bias with the ML-augmented estimates in Figure 5 generated with the Elastic Net predictive model, I observed the same finding regardless of dataset: from 1992 to 2022, there has consistently been no statistically significant difference in change in performance measured by total RAPTOR following a contract year for players coming off a contract year versus those who were not at the 95% level of confidence. (The large error bar in 1999, as mentioned previously, has to do with the nature of the 1999 lockout.)

Next, in Figure 6, we compare estimates that assume that players who did not sign a new contract after their final contract year would have put up nearly impossibly poor numbers the following year should they have signed a new hypothetical contract post-contract year. Even using absolutely worst-case lower bound imputed estimates of player performance to correct for induced bias, the finding remains largely the same: while there is some evidence of players slightly regressing relative to the rest of the league coming off a contract year from 1996 to 1998, there is largely no difference in the change in total RAPTOR relative to non-contract year players, even assuming a worst-case scenario where players are synthetically "forced", in a way, to get worse following their contract year. There is a period from 2015 to 2018

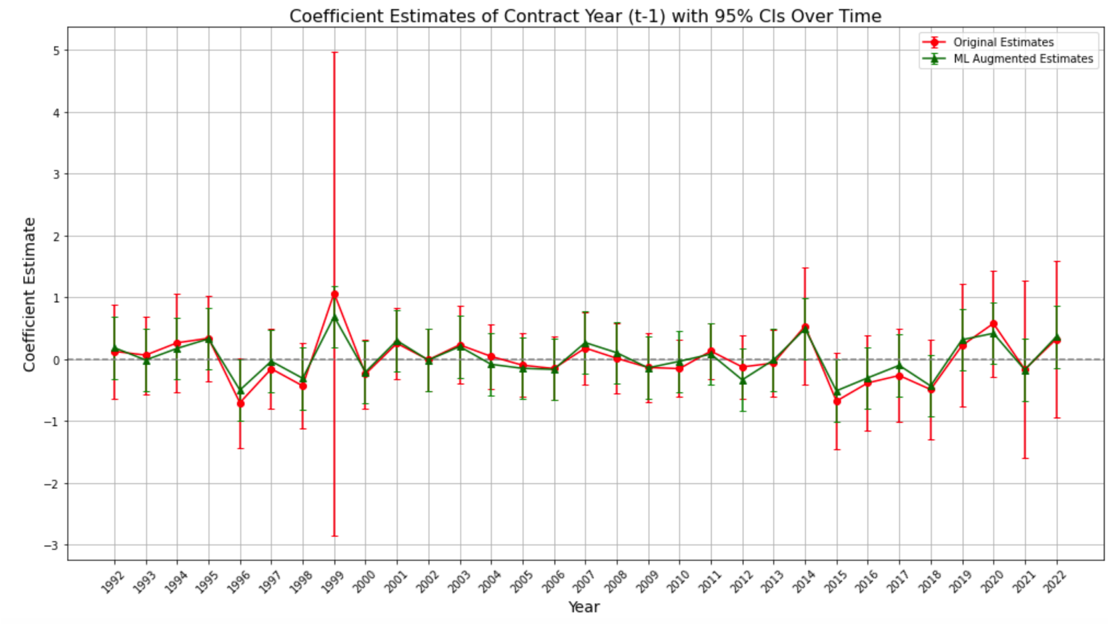

Figure 5: Estimate of Change in Total RAPTOR of Contract Year Players Relative to Non-Contract Year Players in Year Following Contract Year

where the change in performance is negative compared to non-contract year players, but the estimates are still very close to 0 and could change with covariate selection.

Finally, assuming that a player gets no worse than they were in their actual final contract year if they were to sign a hypothetical new contract, the finding remains largely the same in Figure 7, with little difference in change in total RAPTOR for players coming off a contract year relative to non-contract year players. In this scenario, we actually find that in some years, players coming off their contract year on average perform better than their non-contract year counterparts (1994, 2020, 2022). However, note that these estimates are inflated by the nature of how we imputed the hypothetical final year, since assuming that players do not decline whatsoever after their final season is a very strong assumption that does not really hold in the actual NBA data. Even then, the estimates are still close to 0 and could be still prone to noise or sensitive to covariates.

While these plots compare the dynamic association between coming off a contract year and performance measured by the total RAPTOR index, the coefficients of the interaction terms in other models where I changed the dependent variable to the aforementioned statistics does not change the core finding. Regardless of perfor-

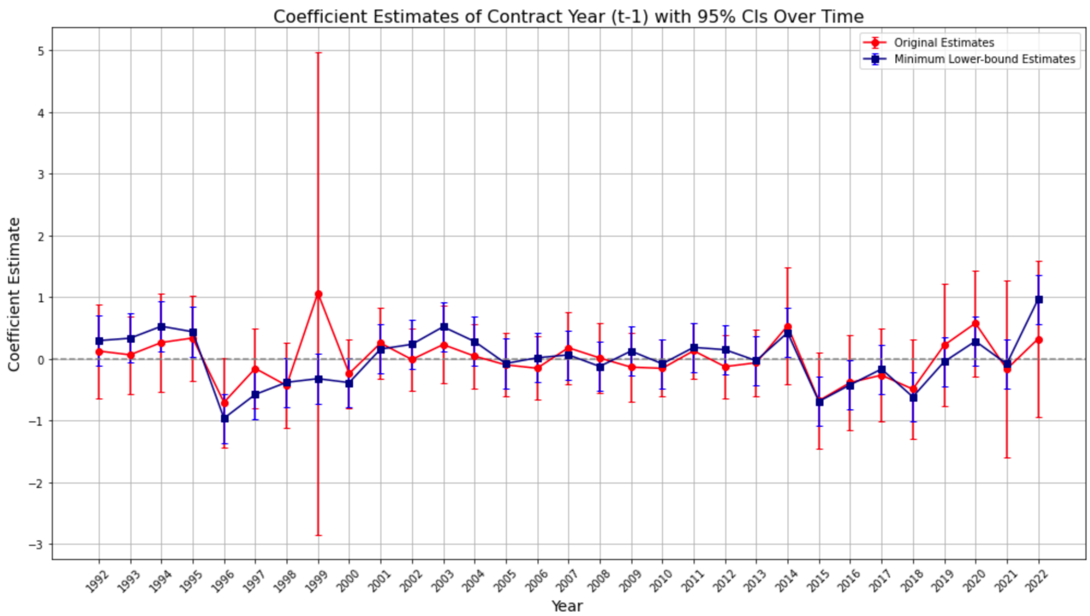

Figure 6: Estimate of Change in Total RAPTOR of Contract Year Players Relative to Non-Contract Year Players in Year Following Contract Year (Original Estimates vs. Worst-Case Lower Bound Imputed Estimates)

mance metric, changes in player performance following a contract year in the NBA are no different than changes in player performance for any other year, suggesting that contract years do not play a huge role in influencing whether players improve or decline year-to-year.

5.3 Addressing Endogeneity with Instrumental Variables

I repeatedly emphasized that the previous findings for differences in mean player performance for players in a contract year versus those not in a contract year is non-causal, as they do not represent the changes in performance of the players in the contract year themselves (rather their average performance relative to non-contract year players); however, even with enough controls or a different model specification, Ryan (2015) points out the inherent endogeneity problem in regression player performance on contract year status. Good players are less likely to be in a contract year in his framework, creating a negative correlation between performance and the probability of being in a contract year.

To get around this endogeneity problem, Ryan regresses a contract year indicator

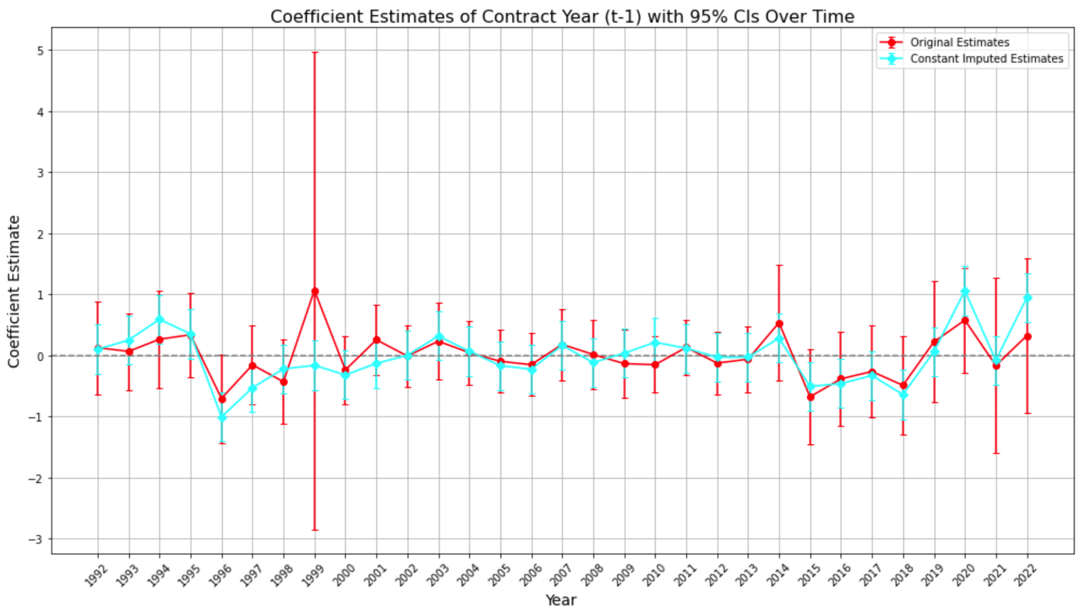

Figure 7: Estimate of Change in Total RAPTOR of Contract Year Players Relative to Non-Contract Year Players in Year Following Contract Year (Original Estimates vs. Imputed Estimates Assuming No Decline in Performance)

on a set of time fixed effects, using a probit link function to generate probabilities of being in a contract year. Ryan claims that year fixed effects are a valid instrument in estimating the second regression in his Two Stage Least Squares estimation, and in doing so, believes he is able to provide a causal estimate of the effect of being in a contract year on player performance. In econometric theory, there are three key assumptions for a valid instrument:

1. Instrument relevance: the instrument needs to be strongly related enough with the endogenous variable of interest (contract year status). Ryan uses the F-statistic from his probit regression of contract year on year fixed effects to demonstrate instrument relevance.

2. Instrument exogeneity: the instrument must be uncorrelated with any unobserved variables that also affect the response variable (player performance metrics). In other words, for instrumental variable Z, $Cov(Z, \epsilon) = 0$, where ϵ is the error term.

3. Exclusion restriction: The instrument must not have any direct effect on the

response variable, in Ryan's case WAR, except through its relationship with the endogenous variable (contract year status).

In this section of the paper, I will attempt to replicate an instrumental variable two-stage least squares regression inspired by the one in Ryan's paper. Unlike Ryan, I opt for a different instrument other than year fixed-effects, which I believe to be insufficient due to the presence of even slight time-variation in total RAPTOR. NBA teams have a certain allotted purse every season that they are allowed to spend on player contracts; this is known as a team's "cap space." I argue this satisfies the exclusion restriction by virtue of the fact that player performance cannot be directly influenced by how much a team spends on its entire roster; there are 15 players on a roster all subject to the same team cap space, and because of the high amount of variation in player performance on a given team, there is not really a feasible way for a single team's cap space to directly influence how a given player on that team performs. However, how much a team has in cap space absolutely affects the terms of contracts given out by teams, which invariably influences contract year status of players on the team.

My first task was to scrape team front office data containing information on their total spending on their rosters, the salary cap in a given season, and the team's cap space in a given season. Due to data availability issues, I was only able to retrieve complete data post-2015, which actually may serve as a good thing. As previously discussed, in 2016 and 2017, the NBA Players' Association famously renegotiated its Collective Bargaining Agreement (CBA), leading to an unprecedented rise in contract sizes, including the "supermax" for the league's best players northward of $200 million over 5 years. As such, post-2016 NBA contracts are fundamentally different from contracts before, so identifying a causal effect of being in a contract year will only be contained to this unique period in NBA history. The data schema for my team management dataset is below.

Column Name	Type	Description
Team	String	NBA team identifier.
Players Active	Integer	Number of active players in the team.
Avg Age Team	Float	Average age of active players in the team.
Total Cap All	Float	Total salary cap used by all players in the team.
Cap Space All	Float	Remaining salary cap space for the team.
Active	Float	Total salary cost of active players.
Active Top 3	Float	Combined salary of the top three earners in the team.
Dead Cap	Float	Salary cost for players no longer active but still impacting the team's salary cap.
Year	Integer	NBA season year.
Win Pct	Float	Winning percentage of the team for the season.

Table 14: Data Schema for NBA Team Salary Cap and Performance

After merging this data with my player data, I proceeded with the following two-stage least squares estimation for contract year status (CY) for player i in year t. Here, $\Phi^{-1}(p)$ represents the probit link function, θ_t are time fixed effects, and $\mathbf{X}$ is a matrix of additional covariates. I opted to use the probit link function not only because it was inspired by the probit function in the first stage least squares in Ryan (2015), but also because it helps constrain the estimated outcomes of the instrument to between 0 and 1 with a directly probabilistic interpretation (since, at the end of the day, we are regressing a binary dependent variable). After obtaining the fitted values for probability of being in a contract year from the first stage, I use those fitted values in the second stage to estimate the effect of contract year status on player performance measured by total RAPTOR.

1. $\Phi^{-1}(\text{CY}_{it}) = \alpha_0 + \alpha_1 \text{Team Cap Space}_t + \theta_t + \gamma^T X + \epsilon_{it}$

2. $\text{Total RAPTOR}_{it} = \beta_0 + \beta_1 \widehat{\Phi^{-1}(\text{CY}_{it})} + \theta_t + \gamma^T X + \epsilon_{it}$

I regress Contract Year onto Cap Space and use the F-statistic as a measure of instrument relevance. (This method is not robust to weak instruments, but provides a useful heuristic.) With an F-statistic 10.51, we have sufficient evidence that our instrument of choice is strongly related enough to contract year status, and proceed with estimating the two stage least squares.

Table 15: TSLS Second Stage Regression Results: Total RAPTOR with Year Fixed-Effects (2016-2022)

Variable	**Coefficient**	**Std. Error**	**95% CI**
Constant	-15.5337***	2.321	[-20.084, -10.984]
Contract_Year_Prob	6.9702***	1.954	[3.140, 10.801]
Possessions (poss)	0.0013***	0.0000766	[0.001, 0.001]
Age	-0.0016	0.018	[-0.037, 0.034]
Games Started (GS)	-0.0035	0.003	[-0.009, 0.002]
Salary	1.4754***	0.193	[1.097, 1.854]
Avg Age Team	0.2164***	0.055	[0.108, 0.325]
Active	4.975×10^{-8}***	1.05×10^{-8}	[2.92×10^{-8}, 7.03×10^{-8}]
Active Top 3	-5.246×10^{-8}***	1.45×10^{-8}	[-8.08×10^{-8}, -2.41×10^{-8}]
Dead Cap	8.67×10^{-8}***	2.27×10^{-8}	[4.22×10^{-8}, 1.31×10^{-7}]
Forward (F)	-0.6420***	0.198	[-1.031, -0.253]
Guard (G)	-0.3554	0.198	[-0.744, 0.033]

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust.

Based on the two-stage least squares estimate, a player in their contract year increases their performance as measured by total RAPTOR by 6.9702 points since the 2016 season, or an increase of 1.3 standard deviations in total RAPTOR just by virtue of being in a contract year. Note that the reason this estimate is much larger than the previous estimates from section 5.1 is that these two coefficients are measuring different things. Section 5.1 was a comparison of the difference in mean performance between contract year players and non-contract year players, while this estimate represents how much players actually increase their production in their contract year, avoiding the initial endogeneity problem. The increase in contract year performance is also robust to choice of performance metric. Whether we measure performance by RAPTOR, WAR, VORP, or BPM, players see a large, statistically significant, causally-identified increase in their performance once they enter a contract year.

Table 16: OLS Regression Results for WAR Total with Year Fixed-Effects (2016-2022)

Variable	**Coefficient**	**Std. Error**	**95% CI**
Constant	-3.4526***	1.154	[-5.714, -1.191]
Contract_Year_Prob	4.2005***	1.076	[2.092, 6.309]
Possessions (poss)	0.0011***	0.0000341	[0.001, 0.001]
Age	-0.0563***	0.006	[-0.068, -0.044]
Games Started (GS)	0.0099***	0.002	[0.006, 0.013]
Salary	1.3121***	0.116	[1.085, 1.540]
Avg Age Team	0.1017***	0.027	[0.049, 0.155]
Active	6.6×10^{-9}	5.34×10^{-9}	$[-3.86 \times 10^{-9}, 1.71 \times 10^{-8}]$
Active Top 3	-2.067×10^{-8}***	7.19×10^{-9}	[-3.48 $\times 10^{-8}$, -6.58 $\times 10^{-9}$]
Dead Cap	2.375×10^{-8}*	1.06×10^{-8}	$[2.93 \times 10^{-9}, 4.46 \times 10^{-8}]$
Forward (F)	-0.5420***	0.081	[-0.701, -0.383]
Guard (G)	-0.5323***	0.082	[-0.694, -0.371]

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust.

Table 17: OLS Regression Results for VORP with Year Fixed-Effects (2016-2022)

Variable	**Coefficient**	**Std. Error**	**95% CI**
Constant	-1.0848**	0.480	[-2.025, -0.145]
Contract_Year_Prob	2.0355***	0.446	[1.161, 2.910]
Possessions (poss)	0.0003***	0.0000145	[0.000, 0.000]
Age	-0.0314***	0.003	[-0.037, -0.026]
Games Started (GS)	0.0057***	0.001	[0.004, 0.007]
Salary	0.6713***	0.048	[0.578, 0.765]
Avg Age Team	0.0387***	0.011	[0.017, 0.061]
Active	3.894×10^{-9}	2.13×10^{-9}	$[-2.89 \times 10^{-10}, 8.08 \times 10^{-9}]$
Active Top 3	-1.109×10^{-8}***	2.94×10^{-9}	[-1.69 $\times 10^{-8}$, -5.32 $\times 10^{-9}$]
Dead Cap	1.551×10^{-8}***	4.16×10^{-9}	$[7.36 \times 10^{-9}, 2.37 \times 10^{-8}]$
Forward (F)	-0.2779***	0.034	[-0.344, -0.212]
Guard (G)	-0.4205***	0.034	[-0.487, -0.354]

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust.

Table 18: OLS Regression Results for BPM with Year Fixed-Effects (2016-2022)

Variable	Coefficient	Std. Error	95% CI
Constant	-10.5109***	1.679	[-13.801, -7.220]
Contract_Year_Prob	6.5029***	1.409	[3.741, 9.265]
Possessions (poss)	0.0012***	0.0000583	[0.0011, 0.0013]
Age	-0.0153	0.014	[-0.042, 0.011]
Games Started (GS)	-0.0062**	0.002	[-0.010, -0.002]
Salary	1.6600***	0.141	[1.384, 1.936]
Avg Age Team	0.1511***	0.046	[0.062, 0.240]
Active	2.342×10^{-8}**	8.13×10^{-9}	[7.48×10^{-9}, 3.94×10^{-8}]
Active Top 3	-4.280×10^{-8}***	1.14×10^{-8}	[-6.51×10^{-8}, -2.05×10^{-8}]
Dead Cap	3.921×10^{-8}**	1.64×10^{-8}	[7.02×10^{-9}, 7.14×10^{-8}]
Forward (F)	-0.4312**	0.154	[-0.733, -0.129]
Guard (G)	-0.9340***	0.155	[-1.237, -0.631]

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust.

6 Conclusion

Based on the available data, there seems to be some strong evidence in favor of a contract year phenomenon, consistent with previous literature on the topic. After controlling for relevant confounders, players in their contract year perform 0.25 to 0.3 points better in the total RAPTOR index compared to players not in their contract year, controlling for season as well. However, these positive differences in performance are driven almost entirely by performance differences on the offensive side of the ball, whereas contract year and non-contract year players are comparable on the defensive side of the ball. The second part of the contract year phenomenon, the idea that players inevitably see a sharp decline in performance following their contract year, does not seem to be supported by the data and is merely a figment of sports urban legend. Compared to players who were not in a contract year in a two-season span, players who exit their contract year do not have a statistically significantly difference in their change in performance, indicating that on average, even after their contract year, players' change in performance does not change much relative to the rest of the league, suggesting that changes in player performance season-to-season are driven by factors other than contract year status in the previous season. Even correcting for induced bias in the original dataset by synthetically generating hypothetical final seasons for players who otherwise "dropped-out" of the league, there was still no significant difference in change in total RAPTOR for contract year players compared to non-contract year players. Finally, while both of these findings are non-causal, there seems to even be some causal evidence in favor of the contract year phenomenon boosting player performance. Using team cap space as an instrument in a two-stage least squares framework, I identify that players in their contract year increase their total RAPTOR performance by nearly 7 points, or over 1.3 SDs, and see similarly sized increases in Wins Above Replacement (WAR), Value Over Replacement Player (VORP), and Box Plus-Minus (BPM). Thus, regardless of performance metric, the contract year phenomenon seems to hold true – following with basic monetary incentives, players likely seek to exert more effort on the court and boost their performance in an attempt to earn more money the following year. An extension of this paper could take a more long-term view of post-contract year performance, since my model of comparison in performance before and after the contract year is restricted only to adjacent years. Additionally, other causal inference methods besides instrumental variables could be used to estimate the causal effect of being in a contract year on performance, and see how the point estimates for the effect compare.

References

Brazda, K.: 2023, Simplifying The Supermax.

Brown, M.: 2015, New CBA Means Better Benefits (And A 45% Pay Bump) For NBA Players.

Jean, N.: 2010, Performance Variation in the NBA: Guaranteed Contracts and the Contract Year Phenomenon.

Reina, C.: 2022, The NBA's Final Seven-Year Man.

Ryan, J.: 2015, Show Me the Money: Examining the Validity of the Contract Year Phenomenon in the NBA.

White, M. and Sheldon, K.: 2013, The contract year syndrome in the NBA and MLB: A classic undermining pattern.

7 Appendix Tables

Table 19: OLS Model Performance Metrics

Metric	RMSE	MAE	Sample SD
GP	0.908196	0.740514	1.0
GS	0.812014	0.637032	1.0
MP	0.721506	0.572416	1.0
FG	0.634983	0.478371	1.0
FGA	0.624953	0.470953	1.0
3P	0.579732	0.402689	1.0
3PA	0.565401	0.390400	1.0
2P	0.642354	0.475362	1.0
2PA	0.636758	0.471397	1.0
FT	0.610134	0.438943	1.0
FTA	0.610216	0.441394	1.0
ORB	0.646661	0.460027	1.0
DRB	0.647979	0.477417	1.0
TRB	0.641524	0.469185	1.0
AST	0.579195	0.390926	1.0
STL	0.683795	0.511399	1.0
BLK	0.652229	0.412391	1.0
TOV	0.666484	0.491992	1.0
PF	0.790710	0.611663	1.0
PTS	0.617010	0.464752	1.0
OWS	0.742676	0.539432	1.0
DWS	0.771794	0.579403	1.0
WS	0.732920	0.545159	1.0
OBPM	0.840555	0.543774	1.0
DBPM	0.920417	0.576571	1.0
BPM	0.855778	0.540945	1.0
VORP	0.660515	0.455440	1.0
raptor_offense	0.896357	0.539189	1.0
raptor_defense	0.962991	0.526981	1.0
raptor_total	0.897685	0.554654	1.0
war_total	0.681652	0.476343	1.0
war_reg_season	0.692842	0.491924	1.0
war_playoffs	0.809081	0.449661	1.0
predator_offense	0.865222	0.538036	1.0
predator_defense	0.891142	0.559879	1.0
predator_total	0.869947	0.565550	1.0
pace_impact	0.952893	0.600623	1.0
Average	**0.738278**	**0.510616**	

Table 20: LASSO Model Performance Metrics

Metric	RMSE	MAE	Sample SD
GP	0.907418	0.746124	1.0
GS	0.810340	0.638865	1.0
MP	0.723902	0.576896	1.0
FG	0.640610	0.482561	1.0
FGA	0.632146	0.475450	1.0
3P	0.586271	0.399290	1.0
3PA	0.573025	0.389204	1.0
2P	0.646169	0.476372	1.0
2PA	0.641610	0.474472	1.0
FT	0.614930	0.438516	1.0
FTA	0.612716	0.440449	1.0
ORB	0.654068	0.456939	1.0
DRB	0.647686	0.474551	1.0
TRB	0.643117	0.465601	1.0
AST	0.586023	0.392529	1.0
STL	0.690297	0.516255	1.0
BLK	0.657527	0.405927	1.0
TOV	0.670915	0.497814	1.0
PF	0.794235	0.611586	1.0
PTS	0.622920	0.469399	1.0
OWS	0.742338	0.538158	1.0
DWS	0.768676	0.577201	1.0
WS	0.731665	0.544641	1.0
OBPM	0.845720	0.546832	1.0
DBPM	0.914330	0.576685	1.0
BPM	0.859955	0.542743	1.0
VORP	0.661791	0.453226	1.0
raptor_offense	0.898279	0.540436	1.0
raptor_defense	0.958203	0.519925	1.0
raptor_total	0.899404	0.556607	1.0
war_total	0.679950	0.471987	1.0
war_reg_season	0.692473	0.489651	1.0
war_playoffs	0.801697	0.431099	1.0
predator_offense	0.869844	0.541119	1.0
predator_defense	0.889061	0.557215	1.0
predator_total	0.872339	0.565050	1.0
pace_impact	0.957074	0.598583	1.0
Average	**0.740506**	**0.510269**	

Table 21: Ridge Model Performance Metrics

Metric	RMSE	MAE	Sample SD
GP	0.907034	0.740094	1.0
GS	0.810791	0.635786	1.0
MP	0.720412	0.571635	1.0
FG	0.634085	0.477475	1.0
FGA	0.624230	0.470450	1.0
3P	0.578998	0.401584	1.0
3PA	0.564290	0.389017	1.0
2P	0.641393	0.473928	1.0
2PA	0.636031	0.470654	1.0
FT	0.609111	0.437634	1.0
FTA	0.609169	0.439855	1.0
ORB	0.645505	0.459063	1.0
DRB	0.646925	0.476770	1.0
TRB	0.640336	0.468225	1.0
AST	0.578279	0.390088	1.0
STL	0.683723	0.511150	1.0
BLK	0.650321	0.410331	1.0
TOV	0.665833	0.491310	1.0
PF	0.789772	0.610858	1.0
PTS	0.616155	0.463619	1.0
OWS	0.741025	0.538155	1.0
DWS	0.769325	0.577220	1.0
WS	0.730909	0.543778	1.0
OBPM	0.839400	0.543172	1.0
DBPM	0.920922	0.576441	1.0
BPM	0.855002	0.540570	1.0
VORP	0.658134	0.453160	1.0
raptor_offense	0.895191	0.538339	1.0
raptor_defense	0.963518	0.526521	1.0
raptor_total	0.898004	0.554576	1.0
war_total	0.679568	0.474152	1.0
war_reg_season	0.691184	0.489977	1.0
war_playoffs	0.805419	0.445905	1.0
predator_offense	0.864106	0.536840	1.0
predator_defense	0.891541	0.559555	1.0
predator_total	0.869898	0.565084	1.0
pace_impact	0.951590	0.598528	1.0
Average	**0.737220**	**0.509500**	

Table 22: Elastic Net Model Performance Metrics

Metric	RMSE	MAE	Sample SD
GP	0.905595	0.742753	1.0
GS	0.808743	0.636020	1.0
MP	0.720991	0.573496	1.0
FG	0.636179	0.479129	1.0
FGA	0.626804	0.471214	1.0
3P	0.582134	0.396361	1.0
3PA	0.568177	0.386582	1.0
2P	0.643334	0.474272	1.0
2PA	0.638550	0.471612	1.0
FT	0.611445	0.435944	1.0
FTA	0.610102	0.438750	1.0
ORB	0.650179	0.456518	1.0
DRB	0.646074	0.473983	1.0
TRB	0.641595	0.465260	1.0
AST	0.581740	0.389166	1.0
STL	0.684914	0.512298	1.0
BLK	0.652721	0.405462	1.0
TOV	0.667800	0.494098	1.0
PF	0.792911	0.610999	1.0
PTS	0.618398	0.465802	1.0
OWS	0.741512	0.537520	1.0
DWS	0.767610	0.575972	1.0
WS	0.730718	0.543655	1.0
OBPM	0.841744	0.543485	1.0
DBPM	0.914154	0.575004	1.0
BPM	0.856172	0.539565	1.0
VORP	0.659079	0.451813	1.0
raptor_offense	0.896220	0.538250	1.0
raptor_defense	0.957769	0.519819	1.0
raptor_total	0.896789	0.553908	1.0
war_total	0.678862	0.471690	1.0
war_reg_season	0.690972	0.488651	1.0
war_playoffs	0.802084	0.433540	1.0
predator_offense	0.867528	0.538422	1.0
predator_defense	0.887779	0.556412	1.0
predator_total	0.869996	0.562877	1.0
pace_impact	0.954457	0.596981	1.0
Average	**0.737887**	**0.508305**	

Table 23: Regression Results for **PTS** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	5.6826*** (0.211)	7.2198*** (0.292)
Contract Year	0.0422 (0.057)	0.0969 (0.058)
Possessions (POSS)	0.0018*** (0.00003)	0.0018*** (0.00003)
Age	-0.1414*** (0.007)	-0.1582*** (0.007)
Games Started (GS)	0.0282*** (0.002)	0.0267*** (0.002)
Salary	1.6340*** (0.047)	1.8333*** (0.052)
Forward (F)	0.8171*** (0.086)	0.8773*** (0.085)
Guard (G)	1.3480*** (0.088)	1.4388*** (0.086)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.737	0.737
Adjusted R-squared	0.736	0.736
F-statistic	707.6	707.6
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 24: Regression Results for **OWS** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-0.7057*** (0.089)	-0.6547*** (0.126)
Contract Year	0.1767*** (0.025)	0.2109*** (0.026)
Possessions (POSS)	0.0007*** (0.00001)	0.0007*** (0.00001)
Age	0.0007 (0.003)	-0.0030 (0.003)
Games Started (GS)	0.0037*** (0.001)	0.0029*** (0.001)
Salary	0.2308*** (0.021)	0.2787*** (0.023)
Forward (F)	-0.0214 (0.034)	-0.0109 (0.035)
Guard (G)	-0.0642 (0.037)	-0.0470 (0.037)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.568	0.571
Adjusted R-squared	0.568	0.570
F-statistic	1240	243.9
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 25: Regression Results for **DWS** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	0.1648*** (0.043)	0.1518** (0.058)
Contract Year	0.0357*** (0.012)	0.0523*** (0.012)
Possessions (POSS)	0.0005*** (0.00001)	0.0005*** (0.00001)
Age	0.0055*** (0.001)	0.0044** (0.001)
Games Started (GS)	0.0004 (0.000)	0.0001 (0.000)
Salary	0.0532*** (0.009)	0.0717*** (0.009)
Forward (F)	-0.3018*** (0.019)	-0.2985*** (0.019)
Guard (G)	-0.6586*** (0.020)	-0.6528*** (0.020)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.689	0.690
Adjusted R-squared	0.688	0.689
F-statistic	2230	432.6
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 26: Regression Results for **WS** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-0.5444*** (0.100)	-0.5103*** (0.142)
Contract Year	0.2118*** (0.028)	0.2623*** (0.030)
Possessions (POSS)	0.0012*** (0.00002)	0.0013*** (0.00002)
Age	0.0063 (0.003)	0.0015 (0.003)
Games Started (GS)	0.0041*** (0.001)	0.0031** (0.001)
Salary	0.2837*** (0.023)	0.3499*** (0.026)
Forward (F)	-0.3219*** (0.042)	-0.3082*** (0.042)
Guard (G)	-0.7219*** (0.043)	-0.6991*** (0.044)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.731	0.734
Adjusted R-squared	0.731	0.733
F-statistic	2556	506.0
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 27: Regression Results for **OBPM** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-4.6438*** (0.206)	-4.2831*** (0.273)
Contract Year	0.1817*** (0.051)	0.1588*** (0.051)
Possessions (POSS)	0.0010*** (0.00003)	0.0010*** (0.00003)
Age	-0.0074 (0.006)	-0.0109 (0.006)
Games Started (GS)	-0.0092*** (0.001)	-0.0100*** (0.001)
Salary	0.5947*** (0.023)	0.6351*** (0.031)
Forward (F)	0.8969*** (0.083)	0.9031*** (0.084)
Guard (G)	1.1552*** (0.084)	1.1685*** (0.084)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.396	0.402
Adjusted R-squared	0.395	0.400
F-statistic	848.2	170.1
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 28: Regression Results for **DBPM** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-1.1672*** (0.105)	-1.3457*** (0.134)
Contract Year	0.0333 (0.028)	0.0107 (0.029)
Possessions (POSS)	0.0002*** (0.00001)	0.0002*** (0.00001)
Age	0.0357*** (0.003)	0.0380*** (0.003)
Games Started (GS)	-0.0039*** (0.001)	-0.0038*** (0.001)
Salary	0.0495*** (0.013)	0.0111 (0.016)
Forward (F)	-0.3560*** (0.046)	-0.3685*** (0.046)
Guard (G)	-0.4637*** (0.047)	-0.4820*** (0.047)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.055	0.057
Adjusted R-squared	0.054	0.055
F-statistic	81.84	16.52
Prob (F-statistic)	8.81e-117	6.05e-105

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 29: Regression Results for **BPM** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-5.8047*** (0.257)	-5.6195*** (0.342)
Contract Year	0.2147*** (0.063)	0.1693** (0.065)
Possessions (POSS)	0.0012*** (0.00003)	0.0012*** (0.00003)
Age	0.0281*** (0.008)	0.0270*** (0.008)
Games Started (GS)	-0.0131*** (0.001)	-0.0138*** (0.001)
Salary	0.6445*** (0.029)	0.6465*** (0.038)
Forward (F)	0.5404*** (0.105)	0.5340*** (0.105)
Guard (G)	0.6914*** (0.106)	0.6864*** (0.106)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.347	0.351
Adjusted R-squared	0.346	0.349
F-statistic	679.3	135.4
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 30: Regression Results for **VORP** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-0.5018*** (0.063)	-0.3119** (0.092)
Contract Year	0.1328*** (0.018)	0.1696*** (0.019)
Possessions (POSS)	0.0004*** (0.00001)	0.0004*** (0.00001)
Age	-0.0053* (0.002)	-0.0093*** (0.002)
Games Started (GS)	0.0042*** (0.001)	0.0037*** (0.001)
Salary	0.2790*** (0.016)	0.3382*** (0.017)
Forward (F)	0.0645* (0.027)	0.0795** (0.027)
Guard (G)	0.1152*** (0.028)	0.1388*** (0.028)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.508	0.514
Adjusted R-squared	0.508	0.513
F-statistic	837.2	168.6
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 31: Regression Results for **RAPTOR Offense** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-4.8622*** (0.273)	-4.8622*** (0.273)
Contract Year	0.1743*** (0.049)	0.1743*** (0.049)
Possessions (POSS)	0.0009*** (0.00002)	0.0009*** (0.00002)
Age	0.0037 (0.006)	0.0037 (0.006)
Games Started (GS)	-0.0105*** (0.001)	-0.0105*** (0.001)
Salary	0.3955*** (0.031)	0.3955*** (0.031)
Forward (F)	1.2143*** (0.075)	1.2143*** (0.075)
Guard (G)	2.1323*** (0.077)	2.1323*** (0.077)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.372	0.372
Adjusted R-squared	0.370	0.370
F-statistic	159.1	159.1
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 32: Regression Results for **RAPTOR Defense** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-0.6105*** (0.155)	-0.7798*** (0.176)
Contract Year	0.0938* (0.042)	0.0758 (0.042)
Possessions (POSS)	0.0002*** (0.00002)	0.0003*** (0.00002)
Age	0.0145** (0.005)	0.0164*** (0.005)
Games Started (GS)	0.0015 (0.001)	0.0016 (0.001)
Salary	0.1208*** (0.021)	0.0814** (0.029)
Forward (F)	-0.7393*** (0.057)	-0.7503*** (0.057)
Guard (G)	-1.1969*** (0.061)	-1.2139*** (0.060)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.085	0.089
Adjusted R-squared	0.085	0.086
F-statistic	146.2	39.51
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.

Table 33: Regression Results for **War Total** with and without Year Fixed Effects

Variable	**(1)**	**(2)**
Constant	-1.6598*** (0.152)	-1.4618*** (0.220)
Contract Year	0.3199*** (0.043)	0.3770*** (0.045)
Possessions (POSS)	0.0012*** (0.00002)	0.0012*** (0.00002)
Age	-0.0013 (0.005)	-0.0084 (0.005)
Games Started (GS)	0.0082*** (0.001)	0.0070*** (0.001)
Salary	0.5119*** (0.037)	0.6067*** (0.041)
Forward (F)	0.1216* (0.063)	0.1433* (0.063)
Guard (G)	0.5097*** (0.066)	0.5450*** (0.066)
Year Fixed Effects	-	Yes
Observations	12491	12491
R-squared	0.561	0.565
Adjusted R-squared	0.561	0.563
F-statistic	1024	206.7
Prob (F-statistic)	0.00	0.00

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are HC2 robust. Year fixed effects are included in the right column.