Does It Pay to Win?: An Exploration of Revenue and Performance in the NBA

Matthew Smith, Joseph Doraid

STAT 139
Final Project
Dec. 2022

Introduction and Motivation

The NBA is one of the world's most important and lucrative sporting competitions, with annual revenues of around $10B annually and millions of global viewers. The sheer size of this spectacle makes the ownership of an NBA franchise a massive privilege shared by only 30 parties worldwide. Beyond the boasting-based incentives that might push one to own an NBA team, there exist clear economic incentives as well. Through advertising sponsorships, broadcasting deals, ticket revenue, and merchandise sales it is estimated that the average profits for the 30 franchises lies in the $100-200M dollar range.[1] Given that all NBA franchise owners are having very lucrative paydays no matter their team's performance, the question arises as to why certain owners choose to invest exorbitant amounts of money in their team's player quality. These investments are particularly mystifying for teams such as the Golden State Warriors, whose owners are expected to pay $483M dollars in salary cap luxury taxes for the 2023-24 season because their player payments supersede the league-allowed limit.[2]

As an attempt to justify these exorbitant investments, we explored the factors that impact year-to-year changes in a team's revenue and franchise value data of the 30 participating franchises from the past 11 years. In doing so, we also hope to explore if strong on-court performance, as quantified by regular season win percentage and the playoff round to which a team reached in any given season, leads to significant increases in value and/or revenue for the owner.

With this project, we are thus looking to answer the following questions:

1. Does on-court team performance, particularly winning a championship, lead to significant increase in team revenue or franchise value in the following seasons?
2. What other factors, beyond on-court team performance, impact changes in a team's revenue and franchise value in the following seasons?

[1] https://www.forbes.com/sites/mikeozanian/2022/10/27/nba-team-values-2022-for-the-first-time-in-two-decades-the-top-spot-goes-to-a-franchise-thats-not-the-knicks-or-lakers/?sh=62025b971cce

[2] https://www.espn.com/nba/story/_/id/34811532/warriors-gm-23-24-tax-see-happens-season

Our initial hypothesis is that strong on-court performance directly leads to future value creation for the owner through revenue and franchise value increases - justifying the significant investments.

Data

For this study, we pooled data from a variety of different sources.

The data regarding the individual teams' on-court performance stems from Basketball Reference, and encodes performance through 5 binary variables representing whether the team reached the Playoffs, Conference Semifinals, Conference Finals, NBA Finals, or won the Championship and a sixth column representing win percentage. Additional data, such as average team age, and average attendance per game were also collected from the same source.

Each franchises' associate metro area population data was collected from The U.S. Census. Per capita income data for these metro areas was collected from the St. Louis FED; we could not find per capita income data for the Toronto metro area, so we imputed it as the median of all other metro area's per capita income in a given year.

Each team's social media footprint was quantified as their number of followers on Facebook and Twitter (Instagram was not selected as it is near impossible to find historical data points for other users), which was collected from Statista.com.

Annual Franchise Value estimates for each team were collected from Forbes.com and annual revenue data (including operating income and ticket revenue) was collected from this RunRepeat article. We adjusted the total and ticket revenue for each team and year by the number of regular season games played. This was necessary because, during the 2011 NBA lockout [3] and 2020 NBA covid bubble,[4] teams played less than the standard 82 regular season games. Thus, in those years, we multiplied total and ticket revenue by the ratio of 82 over the number of regular season games played by each team.

Basic U.S. economic market data (in the form of percent changes in the price of the S&P 500 index over different time horizons) were collected using the Yahoo Finance API.

For each of these metrics, year-by-year data was collected from 2011-2021 for each of the 30 franchises, yielding 330 total observed data points.

Summary of Dataset

The full cleaned data set is set-up as dictated by the below.

[3] https://en.wikipedia.org/wiki/2011_NBA_lockout
[4] https://en.wikipedia.org/wiki/2020_NBA_Bubble

Year: *int representing year of focus.*

Team: *string representing franchise name.*

Ticket: *double representing volume of team ticket revenue for associated year, adjusted for the number of regular season games played in that year (in millions, USD).*

OI: *int representing team operating income for associated year (in millions, USD).*

Population: *int representing population of team's metro area for associated year.*

Income: *int representing per capita income of team's metro area for associated year (in USD)*

Age: *double representing average team age for associated year.*

APG: *int representing average stadium attendance per game for associated year..*

Playoffs: *binary representing if the team made playoffs in the associated year.*

CSF: *binary representing if the team made conference semifinals in the associated year.*

CF: *binary representing if the team made conference finals in the associated year.*

Finals: *binary representing if the team made NBA finals in the associated year.*

Championship: *binary representing if the team won the NBA championship in the associated year.*

WP: *double representing team's win percentage in the associated year.*

Followers: *int representing team's Facebook Likes and Twitter Followers for the associated year (in millions)*

MPC1: *double representing percentage change in price of S&P 500 index from the end of the previous season to the beginning of the current associated season.*

MPC2: *double representing percentage change in price of S&P 500 index from the beginning of the previous season to the beginning of the current associated season.*

MPC3: *double representing percentage change in price of S&P 500 index from the beginning of the current associated calendar year to the beginning of the current associated season.*

Revenue: *int representing total team revenue for associated year, adjusted for the number of regular season games played in that year (in millions, USD).*

Franch_Val: *int representing total team revenue for associated year (in millions, USD).*

For the response variables, instead of using the scraped revenue and franchise value data, we chose to employ the year-over-year percent changes in these values as our response variables given their closer applicability to our stated goal. Thus, two further columns were created regarding revenue - an *FR* column encoding the revenue for the future year (2012 revenue in the row corresponding to 2011), and PC column with that year-on-year percent change. The same process was followed with franchise value (*FVL* and *PC_FVL*).

FR: *int representing total team revenue for following year (in millions, USD).*

PC: *double representing percentage change in revenue from current year to future year.*

FVL: *int representing total team revenue for following year (in millions, USD).*

PC_FVL: *double representing percentage change in franchise value from current year to future year.*

Early Data Analysis (EDA)

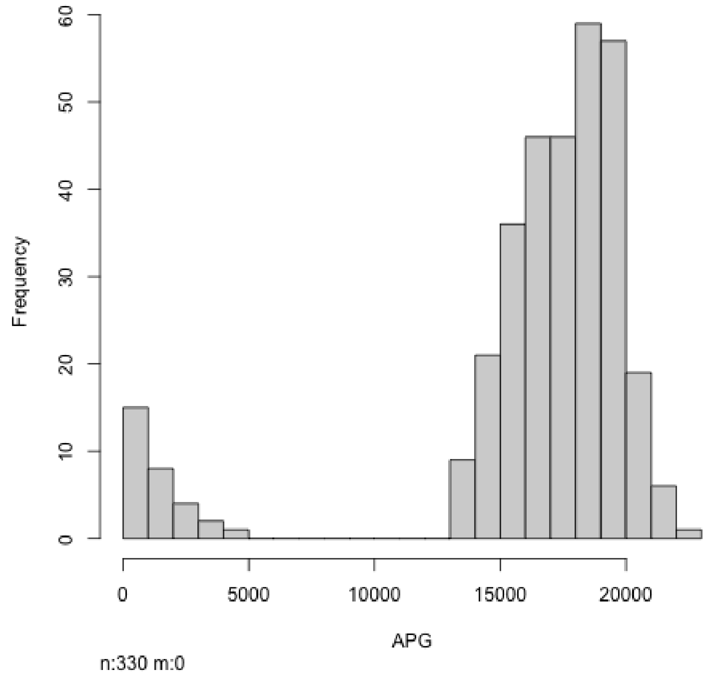
Given that our goal is to explore the effect of on-court team performance on value and revenue creation throughout the NBA, we merged all the aforementioned data sets for the different franchises into a single data set.

Marginal Distribution of Variables

As a first step in our exploration, we inspected the distribution of the continuous numeric values to determine whether they are severely skewed or contain outliers. Based on a visual analysis of the plots in Figures 1-9 of the Appendix, the distributions of six predictors (*Revenue, Ticket, Population, Income, Followers,* and *Franch_Val*) are severely right skewed. Since all of these predictors are strictly non-negative, we may need to perform a log-transformation. Furthermore, two predictors (*OI* and *APG*) have significant outliers. For operating income, the outliers are the Brooklyn Nets in 2014 and 2021. In these years, the Nets paid exorbitant luxury taxes relative to their total revenue.[5,6] To best estimate the true effects of operating income, we may need to remove these outliers. However, as shown in the plot below, we cannot employ the same strategy for the *APG* predictor because a considerable subset of the data follows a disjoint distribution. These outliers are the result of the COVID restriction during the 2020-2021 season. Although *APG* accounts for the decreased number of games played in the 2020-2021 season, it fails to adjust for the health-related capacity restrictions placed on NBA stadiums, which limited game attendance for fans. Some stadiums never allowed any fans to attend games during the season while others limited attendance to a fraction of the total stadium capacity. Given data on the average capacity (whether restricted or not) for each team's stadium in a particular year, we could recalculate *APG* as the average percentage of capacity filled. However, in the absence of this data, we add a binary predictor (denoted *COVID*) indicating whether there were COVID restrictions in a given season; this indicator only has a value of 1 for the 2019-2020 and 2020-2021 NBA seasons. We hypothesize that an interaction term between *COVID* and *APG* will be highly significant.

[5] https://www.sbnation.com/nba/2014/6/30/5857852/brooklyn-nets-financial-reports-losses-lakers-profits

[6] https://www.netsdaily.com/2022/10/28/23428317/forbes-nets-only-nba-team-to-lose-money-over-last-year-but-valuation-rises-9



Histograms of *APG*

Correlation Between Variables

In order to explore the relationships between different predictors and the response variables, we computed the following correlation matrix between all the numeric variables:

	Revenue	Ticket	OI	Population	Income	Age	APG	WP	MPC1	MPC2	MPC3	Followers	Franch_Val
Revenue	1.00	0.60	0.79	0.40	0.29	0.10	0.02	0.06	-0.22	-0.06	0.21	0.63	0.84
Ticket	0.60	1.00	0.67	0.48	-0.11	0.25	0.66	0.11	-0.12	-0.30	-0.04	0.54	0.33
OI	0.79	0.67	1.00	0.30	0.09	-0.05	0.22	-0.02	-0.14	-0.10	0.11	0.55	0.62
Population	0.40	0.48	0.30	1.00	0.38	0.18	0.04	-0.06	0.00	0.01	0.00	0.29	0.43
Income	0.29	-0.11	0.09	0.38	1.00	-0.03	-0.56	0.04	0.05	0.32	0.12	0.23	0.54
Age	0.10	0.25	-0.05	0.18	-0.03	1.00	0.21	0.55	0.01	-0.03	-0.06	0.20	0.04
APG	0.02	0.66	0.22	0.04	-0.56	0.21	1.00	0.16	-0.09	-0.48	-0.17	0.04	-0.28
WP	0.06	0.11	-0.02	-0.06	0.04	0.55	0.16	1.00	0.00	0.00	-0.01	0.07	0.03
MPC1	-0.22	-0.12	-0.14	0.00	0.05	0.01	-0.09	0.00	1.00	0.74	0.76	-0.06	-0.15
MPC2	-0.06	-0.30	-0.10	0.01	0.32	-0.03	-0.48	0.00	0.74	1.00	0.65	0.03	0.10
MPC3	0.21	-0.04	0.11	0.00	0.12	-0.06	-0.17	-0.01	0.76	0.65	1.00	0.11	0.22
Followers	0.63	0.54	0.55	0.29	0.23	0.20	0.04	0.07	-0.06	0.03	0.11	1.00	0.66
Franch_Val	0.84	0.33	0.62	0.43	0.54	0.04	-0.28	0.03	-0.15	0.10	0.22	0.66	1.00

Correlation matrix of numeric variables

As shown above, our first response variable, total revenue, is highly positively correlated with ticket revenue and operating income are highly positively correlated. This is expected as ticket revenue contributes to total revenue and operating income is equal to total revenue minus expenses. Also, as expected, the population (and, to a lesser extent, the average income) of a team's metro area is highly positively correlated with total revenue; population is a significant factor in determining the number of potential fans and thus the number of ticket and merch sales of a team. Interestingly, social media presence is also highly related to revenue. Given revenue-generating sponsorship deals, among other streams that stem from social media presence, it makes sense that such a correlation would exist. Plotted relationships between some of the predictors mentioned and total revenue can be found in Figures 10-12 of The Appendix; these plots further emphasize the strong relationships between variables and highlight some of the outliers discussed previously.

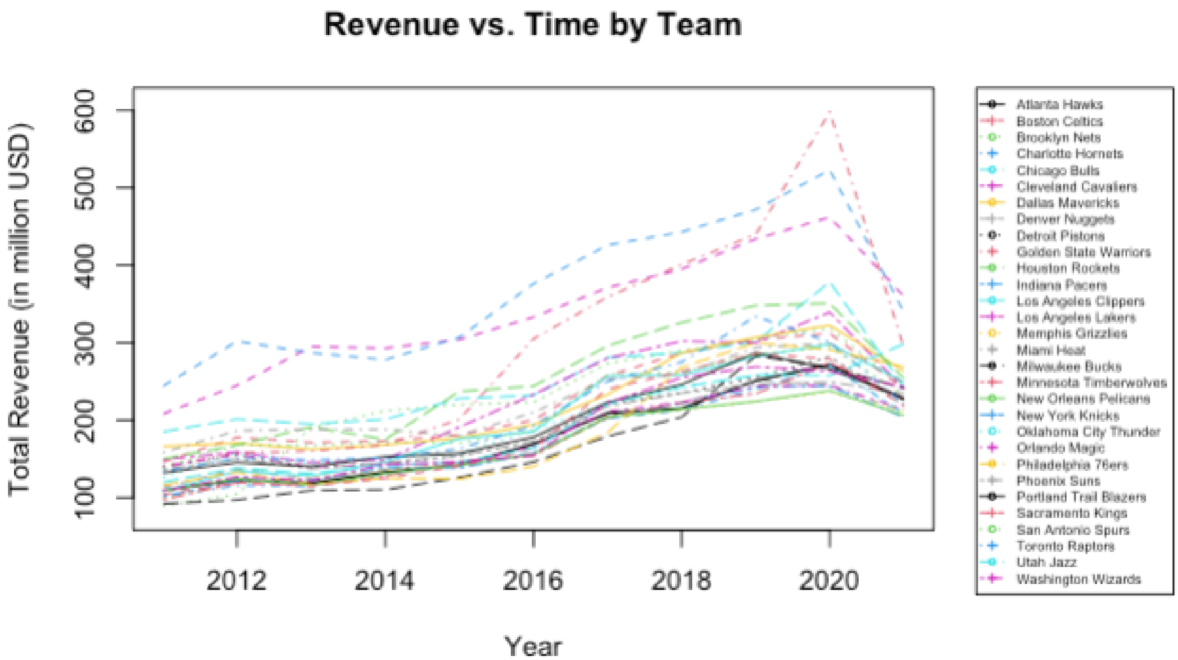
Our second response variable, franchise value, is highly positively correlated with total revenue (as well as operating income and population). This again confirms expectations as a team's revenue stream, expenses (included in operating income), and local market size are major determining factors in their overall evaluation. Furthermore, the correlation between franchise value and social media followers is stronger than that between revenue and followers; the value of a brand more directly affects an organization's overall value than its year-on-year cash flows. Surprisingly, ticket revenue has a weaker relationship with franchise value than total revenue. This may be due to ticket revenue accounting for a very small fraction of the overall, long-term financial success of an NBA franchise. Plotted relationships between some of the predictors mentioned and franchise value can be found in Figures 13-16 of The Appendix.

For both response variables, the win percentage of a team has little to no correlation. This suggests that, without controlling for other factors, the capital at a team's disposal is not associated with winning games. However, this still does not answer the initial question of whether on-court performance affects *future* revenue or franchise value.

A final curious observation worth mentioning is the small correlation between a team's city population and its number of followers on social media, meaning that the vast majority of teams' social media followings and thus global fan bases stem from individuals outside of their local fan bases.

Response Variables Over Time

As a means of exploring overall trends in the response variables across teams, franchise-specific revenue and franchise values were plotted over time. From this visual analysis, three interesting trends surfaced. First, as seen in the graph below, total revenue generally increases steadily over time. This is likely due to the rising popularity of the NBA compounded with inflation. Second, as expected, the Covid-19 pandemic had a strong negative impact on revenue streams. As seen in the graph, almost all teams experience noticeable dips in revenue at the 2020 mark. The only team that successfully recorded revenue growth between 2020 and 2021 was the Utah Jazz, which might be due to the State's limited Covid-restrictions and relatively successful economic pandemic recovery. Third, between each year, most teams observe roughly the same change in revenue. There are a couple of outliers which may be valuable to mention however. For example, the Golden State Warriors notice abnormally steep increases in revenue in the 2014-2018 time frame, directly corresponding to the years in which they won Championships.



Plot of total revenue versus time by team

Unlike revenue, franchise values did not decrease in response to the Covid-19 pandemic. In fact, as shown in Figure 17 of the Appendix, franchise value appears to increase at roughly the same rate from 2020 to 2021 as in prior years.

A deeper dive was also performed visually into the Miami Heat and Golden State Warriors (two of the winningest teams over the past two decades) in order to analyze whether championship-winning seasons led to noticeable increases in value or revenue. As shown in Figures 18 to 21 of the Appendix, in both cases, the associated plots do seem to indicate noticeably faster increases in revenue during seasons following tournament victories. However, year-to-year increases in franchise value appear to occur at roughly the same rate.

Significance of Winning

Finally, we performed further analysis into the relationship between winning a championship and future revenue and franchise value using box-plots and t-tests. For all time lags Δ=1,...,9, we created boxplots of the percent change in revenue and franchise value from year *t* to year *t*+Δ for all years *t* in the dataset, comparing the response between observations when a team won a championship and observations when a team didn't. As shown in these boxplots in Figures 22 to 30 of the Appendix, on average, championship winning teams' actually observe smaller future revenue growth than non-championship teams across all Δ, with the difference being most notable at Δ=6. Performing one-sided t-tests for all Δ and correcting for multiple hypotheses using Bonferroni correction,[7] we find that winning a championship does not lead to a significant

[7] NOTE: we were unable to use more sophisticated correction techniques multiple hypothesis like ANOVA because the number of samples was different for each Δ

increase in revenue growth for any ∆. However, as shown in Figure 40 of the Appendix, we only have 4-5 championship observations for these values of ∆. The low number of samples for one of the classes in the comparison likely skews the results. For franchise value, we do observe an increase in response to winning a championship for $\Delta \leq 5$, as shown in Figures 31-39 of the Appendix. However, for all values of ∆, the t-test fails to reject the null hypothesis.

Results and Methods

Our goal is to find the predictors that most strongly affect the value creation of an NBA franchise, with the hypothesis that a team's on-court performance will be one of the most significant indicators. In order to explore this, we ran six different models: a simple intercept-only linear model, a baseline linear model including all predictors, a full linear model with all predictors including interaction terms, a stepwise model, and two mixed effects models.

These models were run separately to predict annual percentage change in total revenue (*PC*) and annual percentage change in franchise value (*PC_FVL*) as the response variable. We chose year-to-year percent change rather than larger time horizons because it maximizes the number of observations available to fit our models.

As shown in Figure 42 of the Appendix, *PC_FVL* is heavily right skewed, so we apply a log-transformation. We also apply the log-transformation to the six heavily right-skewed predictors (*Revenue, Ticket, Population, Income, Followers,* and *Franch_Val*) identified in the EDA section. We do not apply a transformation to *PC* as it is still roughly symmetric, even though it has long tails as shown in Figure 41 of the Appendix.

We chose to use Bayesian Information Criterion (BIC) as our evaluation criterion for selecting the most appropriate model.

In terms of data preparation, for each run-through three NBA teams were selected at random as the test set - thus, the models were trained on the other 27 franchises.

Baseline Model

First, we fit baseline models for each response variable to:

1. Identify outliers and violations of assumptions
2. Identify significant predictor variables in this baseline model
3. Obtain a benchmark to compare with other models

We fit a standard OLS model with terms for all the numeric variables; we do not include the categorical variable *Team* in these models.

In Figures 43 and 44 of the Appendix, we visually investigate whether each of the following assumptions of OLS hold for the *PC* baseline model:

- Linearity: in the plot of the residual versus fitted values (Figure 43), we see that the trendline hovers around zero, suggesting that linearity is satisfied.
- Equal Variance: in the plot of the residual versus fitted values (Figure 43), we see that the variance for smaller fitted values (i.e. less than 1) is much lower than that of larger values. This is a result of the noticeable change in revenue growth in the NBA during the Covid-19 pandemic. As previously mentioned, we hope that an interaction term between *APG* and *COVID* will correct for this violation.
- Normality: in the Q-Q plot (Figure 44), we see that the observed distribution of data roughly follows the theoretical quantiles of the normal distribution. However, there are few outliers which significantly diverge significantly from these quantiles; the outliers were the 2012 Brooklyn Nets and 2020 Utah Jazz, the unique circumstances of which are discussed earlier in this paper. We remove both the outliers from the training set to correct for this violation of assumptions.

After correcting for assumption violations, the *PC* baseline model achieves an R^2 of 0.59 and BIC of -407.27 on the training set. Furthermore, the following predictors are identified as significant under this model: *Revenue*, *OI*, *Age*, *WP*, *Franch_Val*, and *COVID* (see Figure 45 of the Appendix). Notably, *WP* has a positive coefficient associated with it.

In Figures 46 and 47 of the Appendix, we visually investigate whether each of the following assumptions of OLS hold for the *PC_FVL* baseline model:

- Linearity: in the plot of the residual versus fitted values (Figure 46), we see that the trendline has a slight quadratic relationship, suggesting that linearity is violated. However, the trendline is considerably skewed by two outliers.
- Equal Variance: in the plot of the residual versus fitted values (Figure 46), we see that variance is roughly the same across all fitted values.
- Normality: in the Q-Q plot (Figure 47), we see that the observed distribution of data roughly follows the theoretical quantiles of the normal distribution. However, there are few outliers which significantly diverge significantly from these quantiles; the outliers were the 2012 Sacramento Kings and 2014 Los Angeles Clippers. In the case of the Kings, the owners of the team were on the verge of bankruptcy and failed multiple times to relocate the team before selling it at a discounted price.[8] We remove both the outliers from the training set to correct for the violation of normality and linearity.

After correcting for assumption violations, the *PC_FVL* baseline model achieves an R^2 of 0.58 and BIC of -370.31 on the training set. Furthermore, the following predictors are identified as significant under this model: *Population, Income, MPC1, MPC2, MPC3, Followers, Franch_Val,* and *COVID* (see Figure 48 of the Appendix).

For all future models considered, the correction of assumption violations still apply. However, unlike this section, the p-vals for coefficient estimates can no longer be considered meaningful because testing coefficients under multiple models can lead to inflated significance.

[8] https://en.wikipedia.org/wiki/Failed_relocation_of_the_Sacramento_Kings

Intercept Model

A simple intercept-only OLS model was also created as a benchmark with which future iterations of more complex models could be compared, particularly for use in the development of the stepwise model. The reported BIC when run for *PC* was -278.88 (see Figure 49 in the Appendix) For *PC_FVL*, the BIC was -249.84 (see Figure 50 in the Appendix)

Full Interaction Model

Next, we consider a linear regression model containing all numeric predictors and their two-way interaction terms. We do not include the *Team* predictor since this would create more parameters than observations in the training set.

For the response variable *PC*, despite producing a high R^2 of 0.93, the full interaction model receives a much larger BIC of 7.46 (see Figure 51 of the Appendix). Similarly, for *PC_FVL*, the full interaction model has a reported R^2 of 0.95 and BIC of -75.17 (see Figure 52 of the Appendix). We consider this model as an upper bound in performance that severely overfits the training data and does not actually approximate the ground truth model for changes in revenue and franchise value.

Stepwise Model

With both lower and upper bound models having been developed, a stepwise model was then trained in order to better approximate the predictors and interaction terms in the ground truth model for revenue and franchise value. For both response variables, we ran combined stepwise model selection using the intercept-only model as the lower bound and initial model considered and the full interaction model as the upper bound.

For the response variable *PC*, the final stepwise model includes just one predictor, *COVID*, and no interaction terms. The model has a reported R^2 of 0.40 and BIC of -411.26 (see Figure 53 of the Appendix).

For the response variable *PC_FVL*, the final stepwise model includes the following predictors: *MPC1, MPC2, MPC3, COVID, Income,* and *WP*. It also includes two interaction terms: *MPC2:MPC1* and *MPC1:MPC3*. The model has a reported R^2 of 0.73 and BIC of -555.14 (see Figure 54 of the Appendix). Notably, *WP* has a positive coefficient associated with it.

Mixed-Effects Models

Finally, we decided to run a mixed effects model conditioning on the specific year. One can consider each team's percent change in revenue or franchise value as being drawn from a new random distribution every year, justifying the selection of this model. This interpretation is supported visually in the plot of each of these values over time across teams in the EDA section.

First, we consider the mixed effects model with only a random intercept conditional on year. For the response variable *PC*, this model yields a conditional R^2 of 0.66 and a BIC of -514.38 (see Figure 55 of the Appendix). For *PC_FVL*, this model has a reported conditional R^2 of 0.73 and BIC of -544.27 (see Figure 56 of the Appendix). For mixed effects models, the conditional R^2 is the proportion of variance in the response variable explained by fixed and random effects.[9]

Second, we consider the mixed effects model which includes the random intercept conditioned on year and any fixed effects in the stepwise model which vary in a given year across teams. Including variables like *MPC1*, which are constant in a given year, would be redundant. Since the stepwise model for *PC* only includes one predictor, which is specified entirely by year, we do not add any fixed effects. However, for *PC_FVL*, we add two fixed effects: *WP* and *Income*. For *PC_FVL*, this model has a reported conditional R^2 of 0.77 and BIC of -570.00 (see Figure 57 of the Appendix). Notably, *WP* has a positive coefficient associated with it.

Summary of Results

Below, we provide a summary of the performance metrics of all the models considered, with the highest performing in each category bolded:

Model	R^2, *PC*	BIC, *PC*	R^2, *PC_FVL*	BIC, *PC_FVL*
Baseline	0.59	-407.27	0.58	-370.31
Intercept-Only	N/A	-278.88	N/A	-249.84
Full Interaction	**0.93**	7.46	**0.95**	-75.17
Stepwise	0.40	-411.26	0.73	-555.14
Mixed Effects, Intercept Only	0.66	**-514.38**	0.73	-544.27
Mixed Effects, Random Intercept + Fixed Effects	N/A	N/A	0.77	**-570.00**

Validation

We tested the highest performing BIC models on the validation set. For *PC*, the intercept-only mixed effects model had a reported R^2 of 0.84 on the validation set. For *PC_FVL*, the random intercept + fixed effects model had a reported R^2 of 0.74. Given that neither of the models experience a significant decrease in R^2 on the validation set, these results confirm that the

[9] https://jonlefcheck.net/2013/03/13/r2-for-linear-mixed-effects-models/

selected models generalize to unseen data and are thus appropriate for the response variables of interest.

Interpretation

In the initial baseline models, the t-tests on the significance of coefficients gave strong, early indications of which factors truly influence changes in revenue and franchise value. For franchise value, it appeared that the most significant factors were determined off the court (i.e. not related to performance) and out of the front office (i.e. not related to a team's spending habits or brand), from changes in the overall US stock market to the income per capita of a team's associated metro area. Unsurprisingly, a teams' brand presence, popularity, and "star power," as estimated by a team's social media following, also significantly impacted changes in franchise value under the baseline model. While off-court factors like Covid-19 restrictions were also significant for revenue changes, a team's on-court makeup and performance in the form of average age of player and win percentage also played a role, with the model predicting that younger and more winning teams would experience higher percent changes in revenue.

The stepwise models for both response variables challenged these initial conclusions. For revenue, only considering whether there were Covid-19 restrictions in place yielded a lower BIC than the baseline, suggesting that it was a more appropriate representation of the ground truth model. Thus, under this new model, a team's performance played no role in changes in revenue. In contrast, for franchise value, the stepwise model incorporated win percentage in addition to many of the aforementioned macro-level market factors. Like the baseline model for revenue, it suggests that win percentage increases the change in franchise value.

The stepwise models also rejected some of our initial hypotheses about which interaction terms would be significant. In particular, we expected many of the predictors which are dependent on time, like revenue, to appear as interaction terms. Similarly, we expected the inclusion of an interaction term between *APG* and *COVID*, as the former variable decreased significantly under Covid-19 restrictions. Neither hypothesis turned out to be the case.

Ultimately, variations of the mixed effect models with intercepts conditional on year proved to be the most appropriate fit, confirming our intuition that the year-to-year changes for each team are sampled from the same conditional distribution. For revenue, the intercept-only mixed effects model vastly outperforms the other models considered in BIC. This suggests that some other league-wide, year-to-year variation, outside of the predictors considered, explains year-to-year changes in revenue. One drawback to the mixed effects models discussed is that it assumes that we can perfectly estimate the distribution of the percent change in revenue each year. This is a very strong assumption, so it is worth investigating how well we can actually estimate these distributions and predict them in future years. For franchise value, the combined mixed effects model (i.e. random intercept + fixed effects) only marginally outperforms the stepwise model in R^2 and BIC. This suggests that the year-to-year variations fit by the random intercepts are mostly explained by the previously considered market-wide factors. Additionally, the improvement in BIC from adding the fixed effects drives home the significant role of win percentage in increasing future franchise value.

Discussion and Conclusion

We set out to explore whether an NBA team's on-court performance plays a significant role in predicting its monetary value creation, as quantified by the year-on-year percentage changes in revenue and franchise value. Our investigation seems to show that, when considering all 30 NBA teams, this is partially the case. Although whether or not a team wins the Championship does not necessarily hold a strong correlation to value creation, team success as dictated by win percentage (*WP*) definitely does. This relationship definitely isn't as strong as we would have predicted, however, given the exorbitant sums of money franchise owners invest in their teams' quality. In fact, our analysis seems to show that owners are quite powerless when it comes to increasing their asset's value, with both revenue and franchise value shifts being more directly related to macroeconomic trends than to performance. Overall, this is not a surprise given the consumer-based business models of all these franchises. If the economy is doing poorly, individuals are less likely to spend their money on products such as game tickets, streaming services, and team merchandise, directly impacting revenue streams, and thus franchise values. This correlation was made painfully obvious by the Covid pandemic, an unprecedented period that dramatically affected the US economy and thus consumer spending, affecting revenue and franchise value in exponentially greater ways than any on-court performance might.

There are, of course, important limitations to our study. A first one concerns our period of focus, which includes the aforementioned unprecedented Covid pandemic. Using this time frame might be problematic given the violent and generally unrepresentative effects the pandemic had on not only the economy, but the world as a whole. It essentially invalidated the generalizability of the last two years (~20% of our sample size). It is very unlikely that the revenue of the Los Angeles Lakers or the Milwaukee Bucks would have been similarly affected by their respective 2020 and 2021 Championship victories had there not been a pandemic. For the Lakers in particular, whose entire Championship run was achieved within the NBA's Covid-safety bubble, the franchise would have seen a fraction of the expected revenue increases had they won during a regular season (e.g. no ticket-sale revenue). Thus, the percentage changes in revenue and franchise value faced by these teams following their 2020 and 2021 seasons definitely are not generalizable to the vast majority of other NBA seasons, perhaps negatively impacting our model's predictiveness. Another potential issue concerns our use of the single year-on-year percentage change in revenue and franchise value, without a consideration of longer term shifts. It is very possible that franchises see the effects of successful on-court performances (Championship victories) at a lagged rate, as it takes them time to expand their brand and capitalize on their feat. Perhaps, by zooming out to percentage changes over longer periods, we may be able to more accurately predict value creation through tournament success. However, our ability to increase our lag window is severely limited by the small sample size of championship teams (1 per year).

Overall, our analysis does encourage a certain level of investment from NBA franchise owners into their team quality as a means of creating monetary value for themselves, especially in the form of franchise value. With that said, it also reminds us that factors beyond the control of any owner ultimately play the strongest effects, with changes in revenue and franchise value ultimately being at the mercy of macroeconomic performance.

Appendix

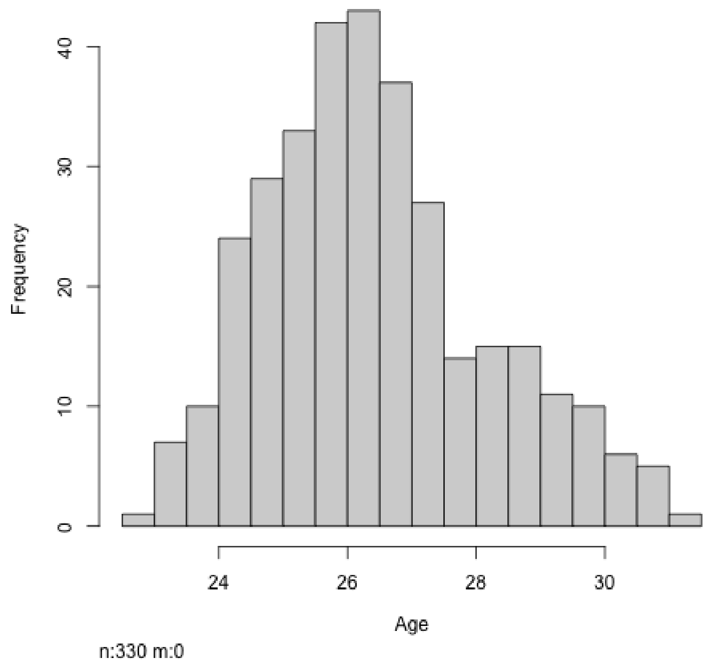


Figure 1: Histogram of *Age*

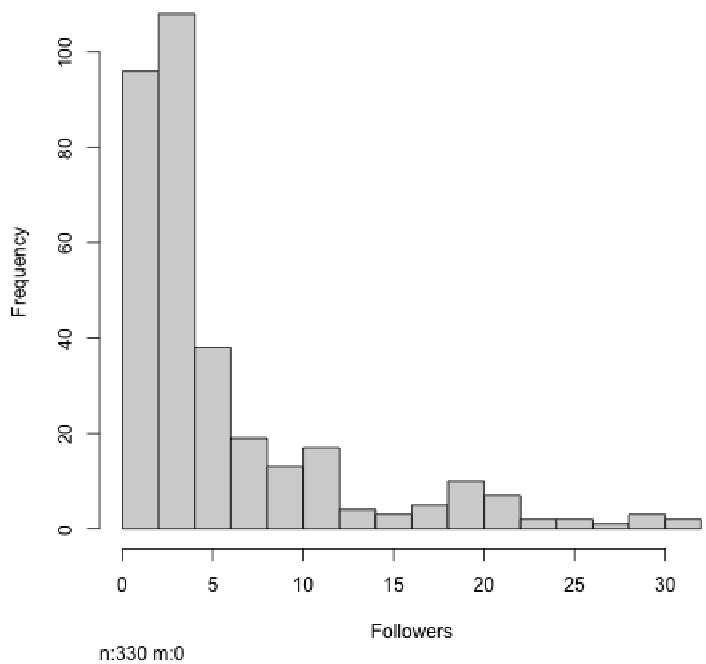


Figure 2: Histogram of *Followers*

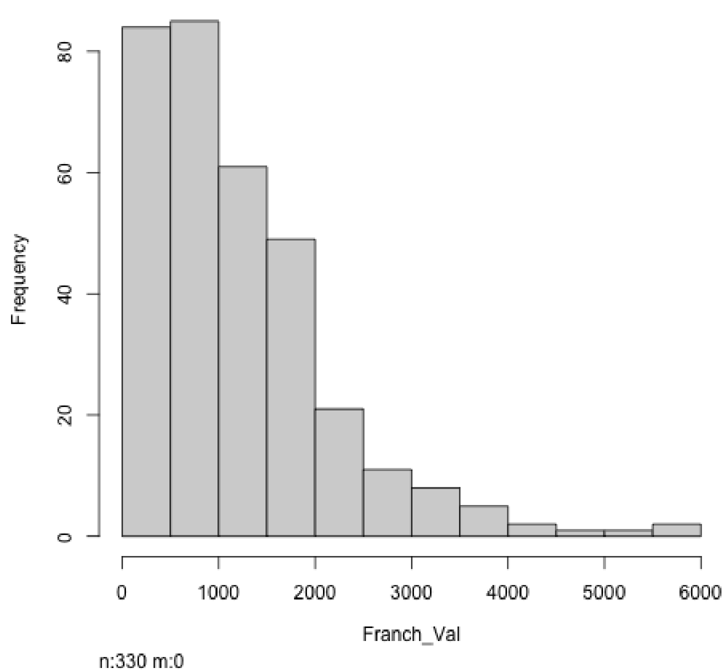


Figure 3: Histogram of *Franch_Val*

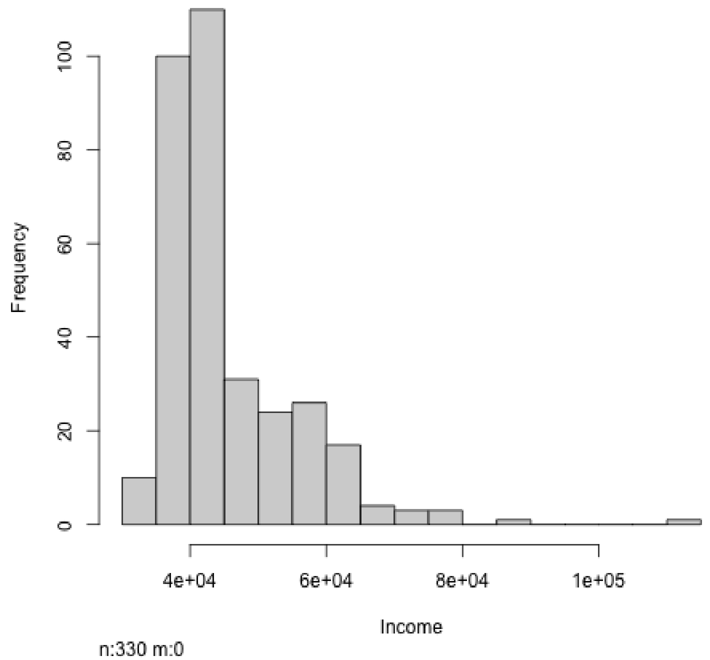


Figure 4: Histogram of *Income*

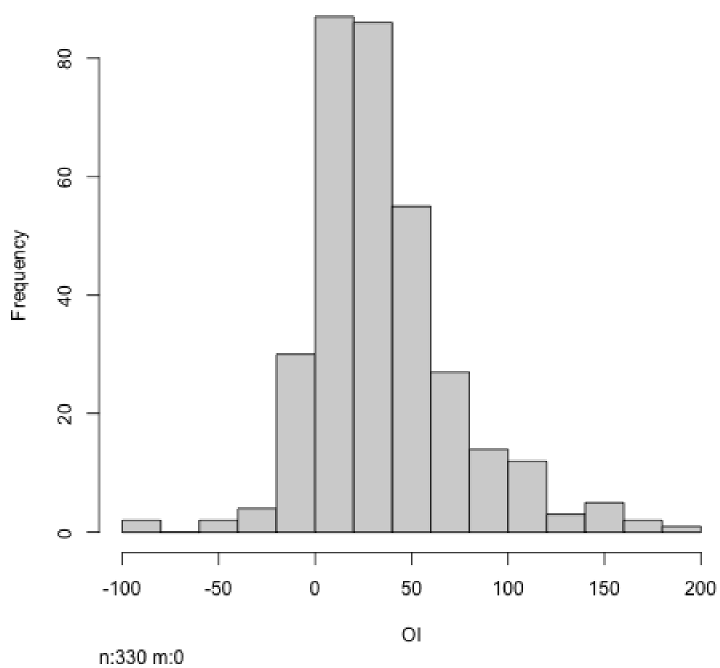


Figure 5: Histogram of *OI*

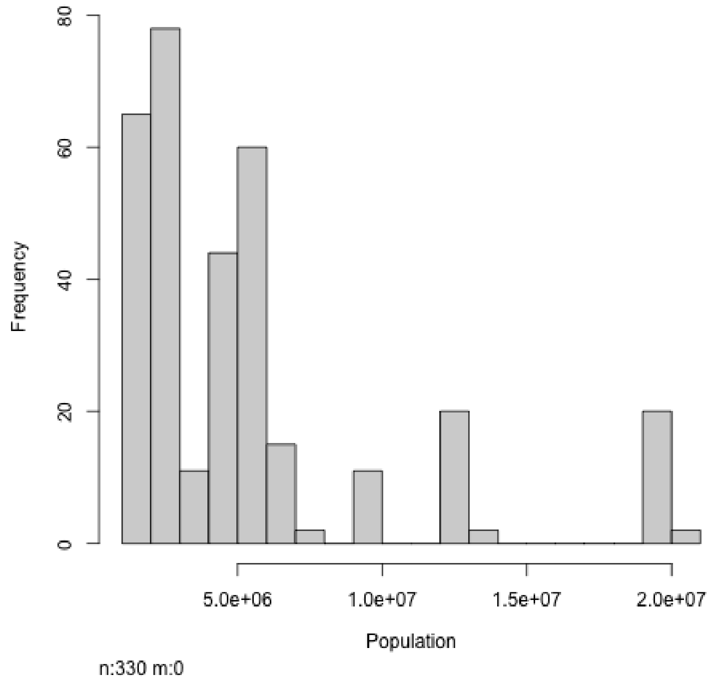


Figure 6: Histogram of *Population*

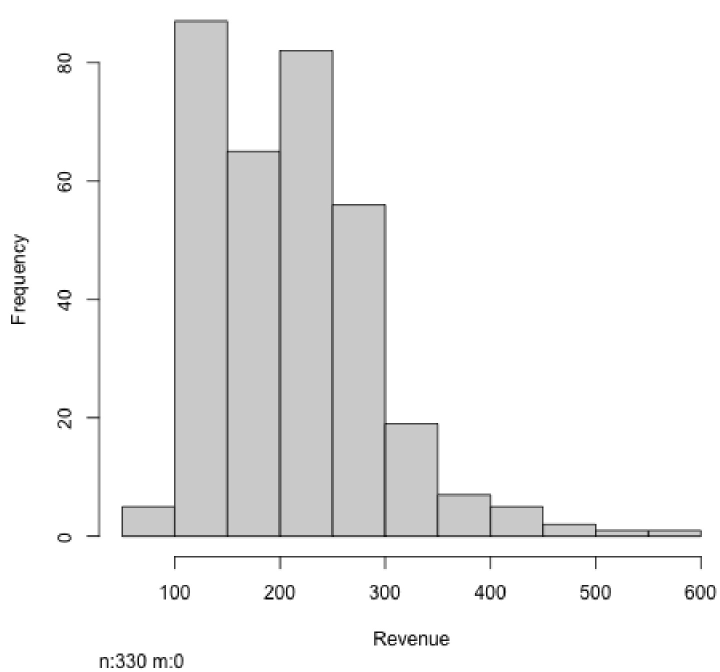


Figure 7: Histogram of *Revenue*

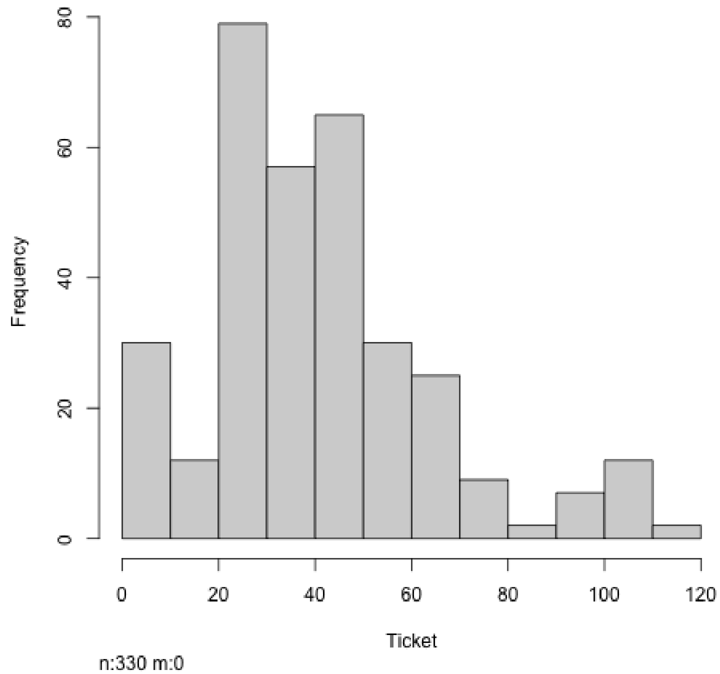


Figure 8: Histogram of *Ticket*

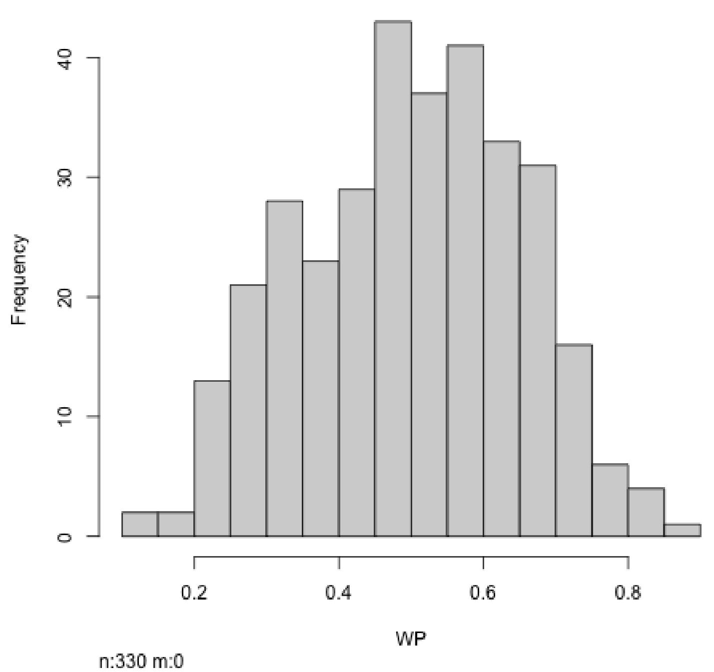


Figure 9: Histogram of *WP*

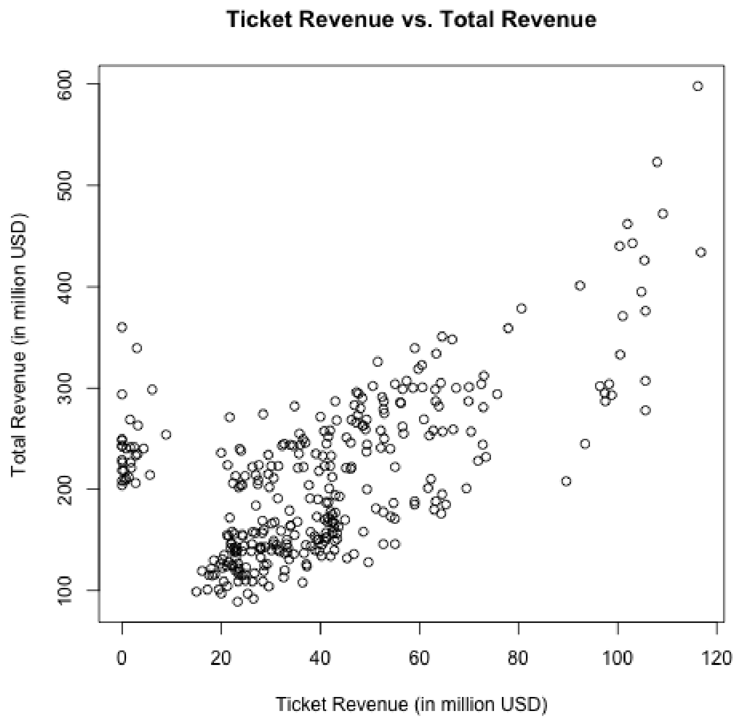


Figure 10: Scatterplot of ticket revenue versus total revenue

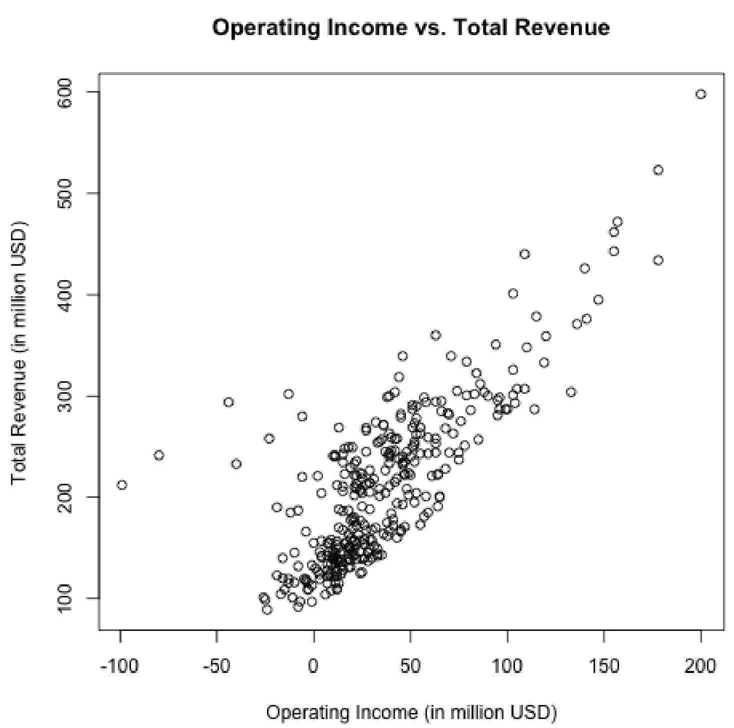


Figure 11: Scatterplot of operating income versus total revenue

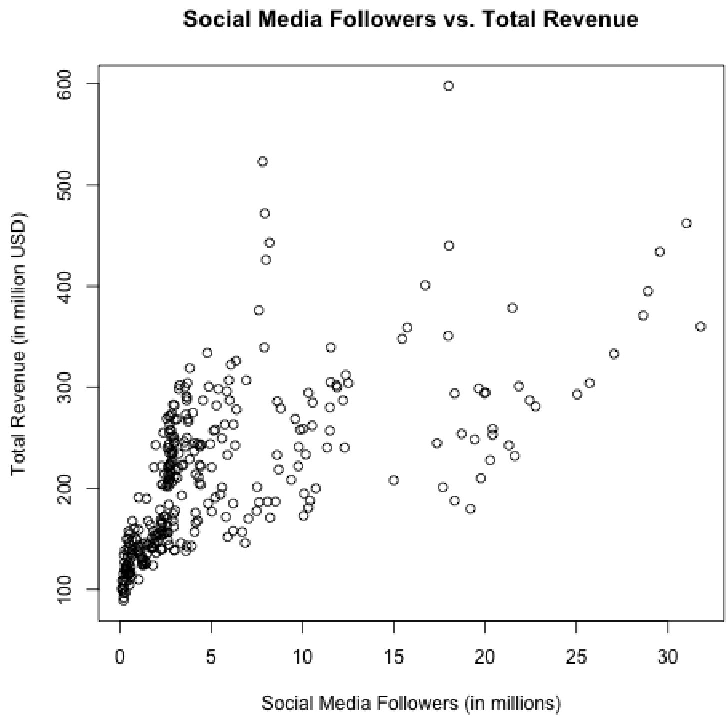


Figure 12: Scatterplot of social media followers versus total revenue

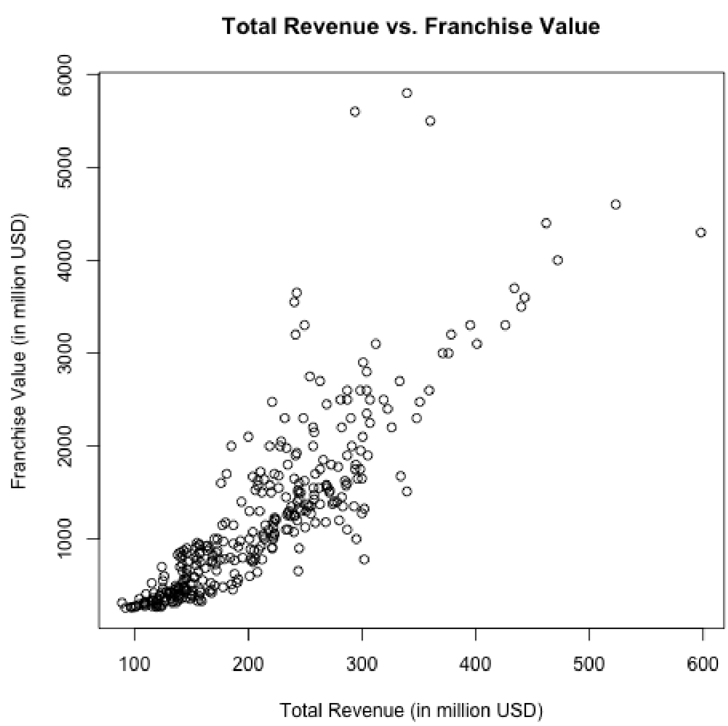


Figure 13: Scatterplot of total revenue versus franchise value

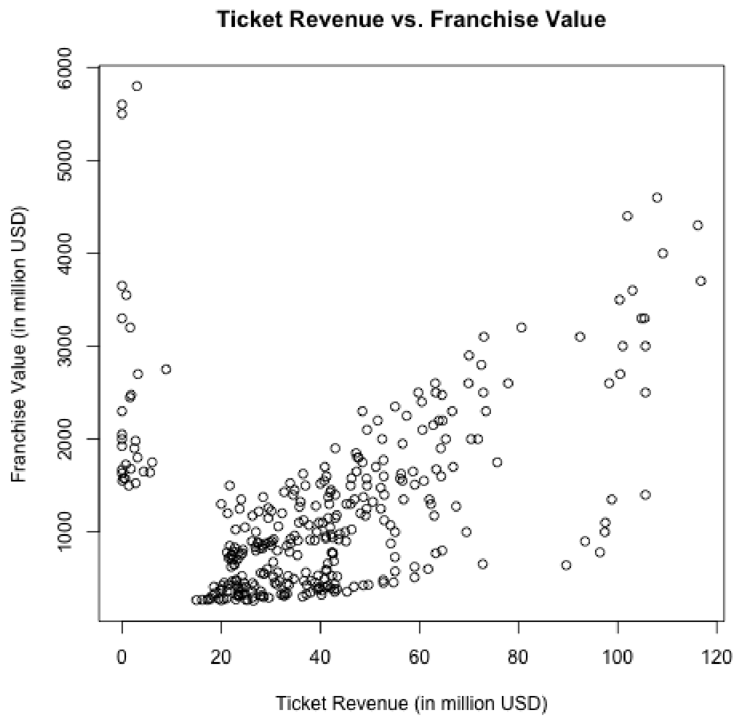


Figure 14: Scatterplot of ticket revenue versus franchise value

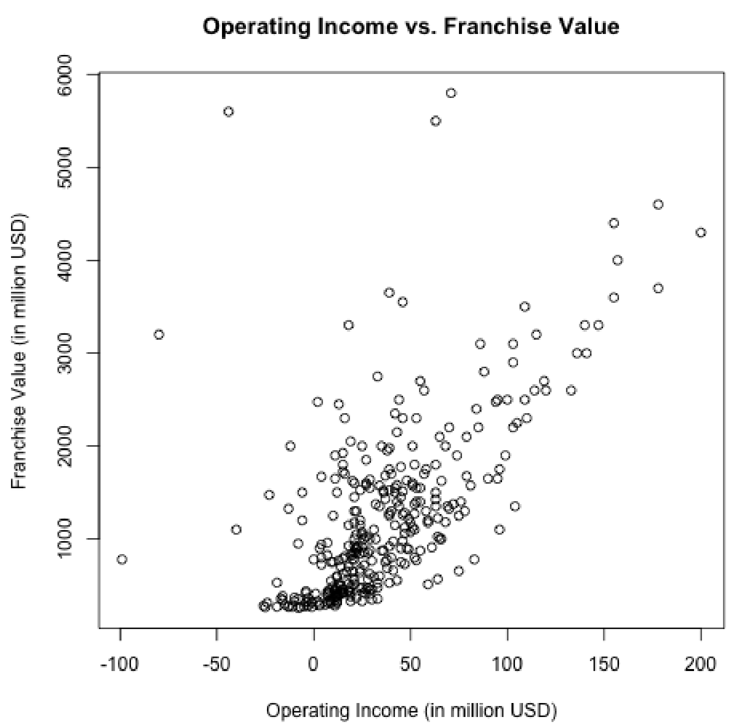


Figure 15: Scatterplot of operating income versus franchise value

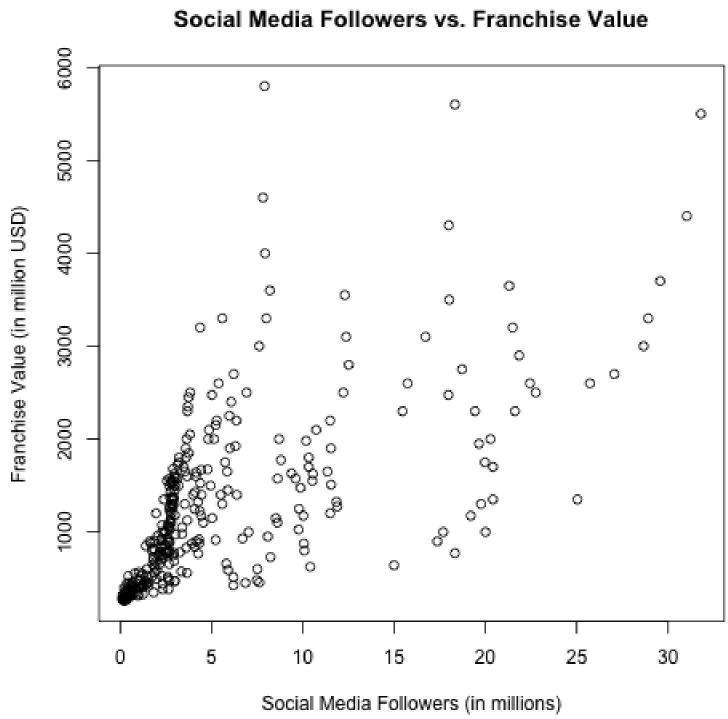


Figure 16: Scatterplot of social media followers versus franchise value

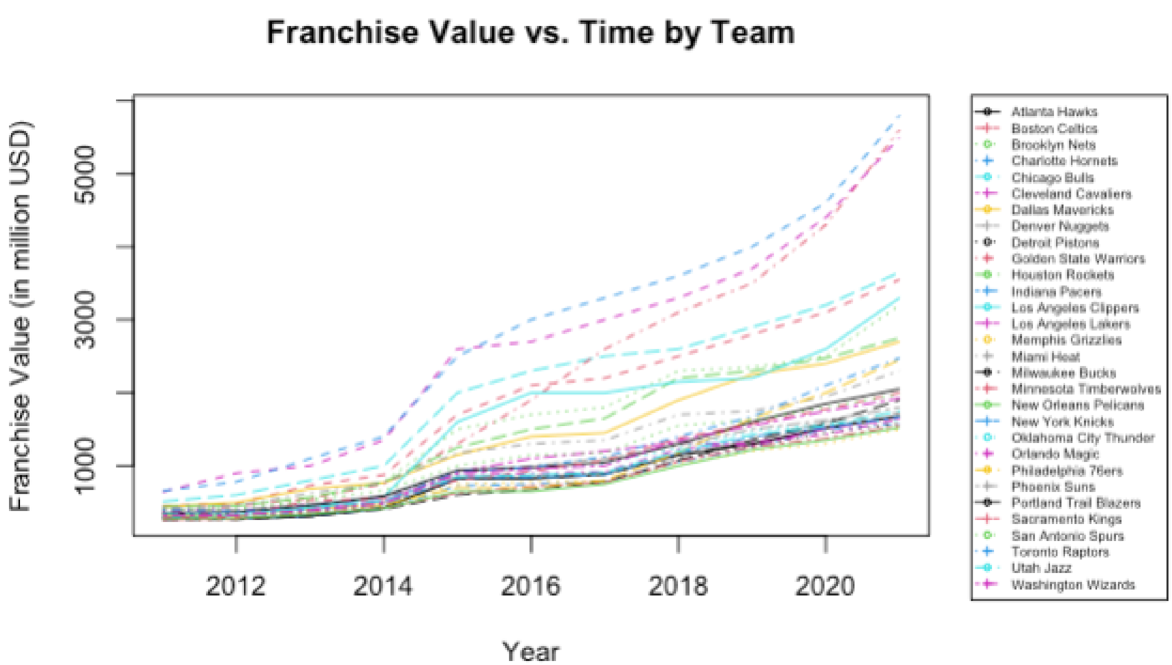


Figure 17: Plot of franchise value versus time by team

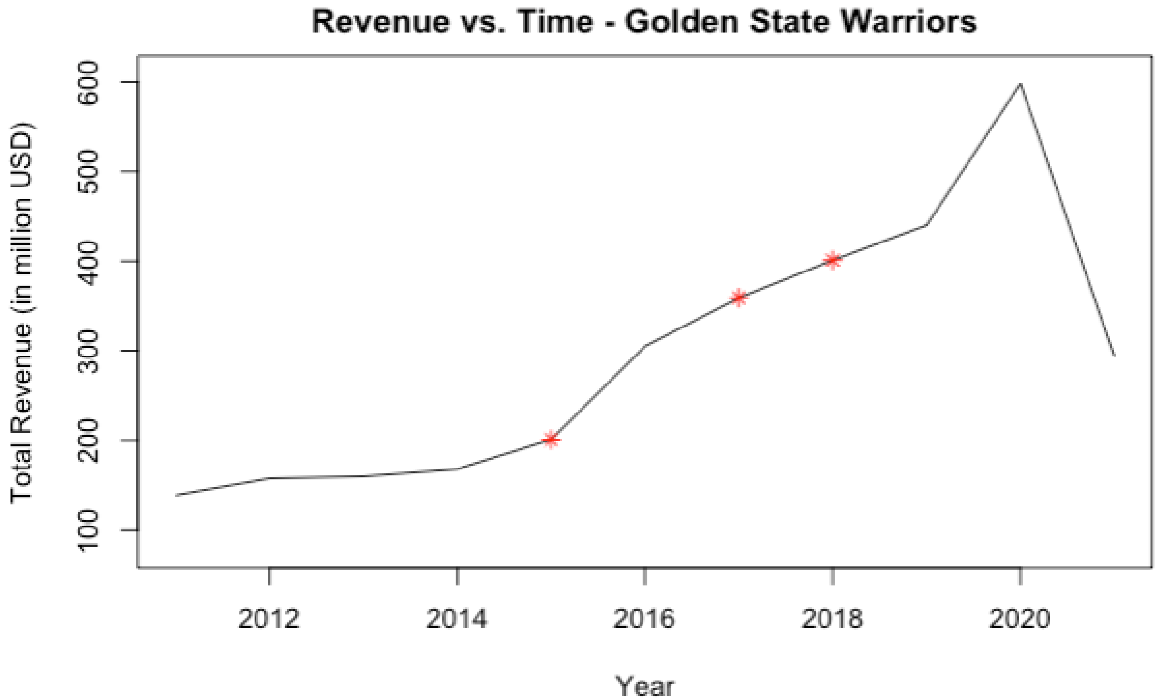


Figure 18: Plot of revenue versus time for the Golden State Warriors. The red stars indicate years when the Golden State Warriors won a championship.

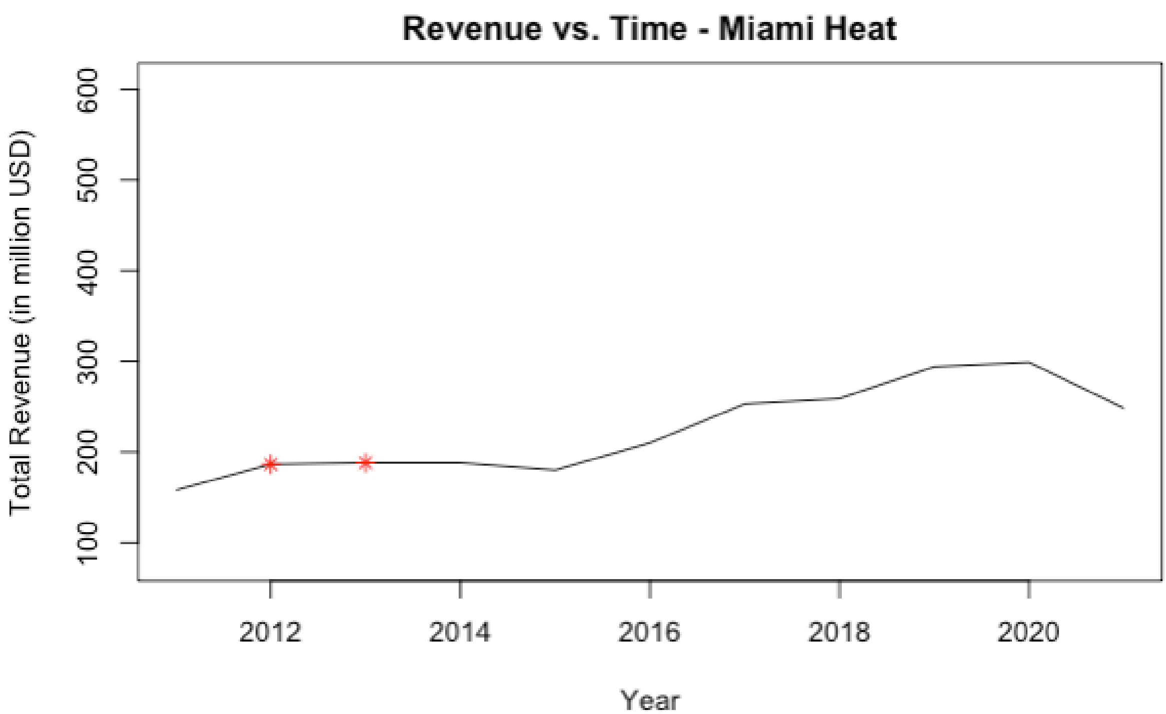


Figure 19: Plot of revenue versus time for the Miami Heat. The red stars indicate years when the Miami Heat won a championship.

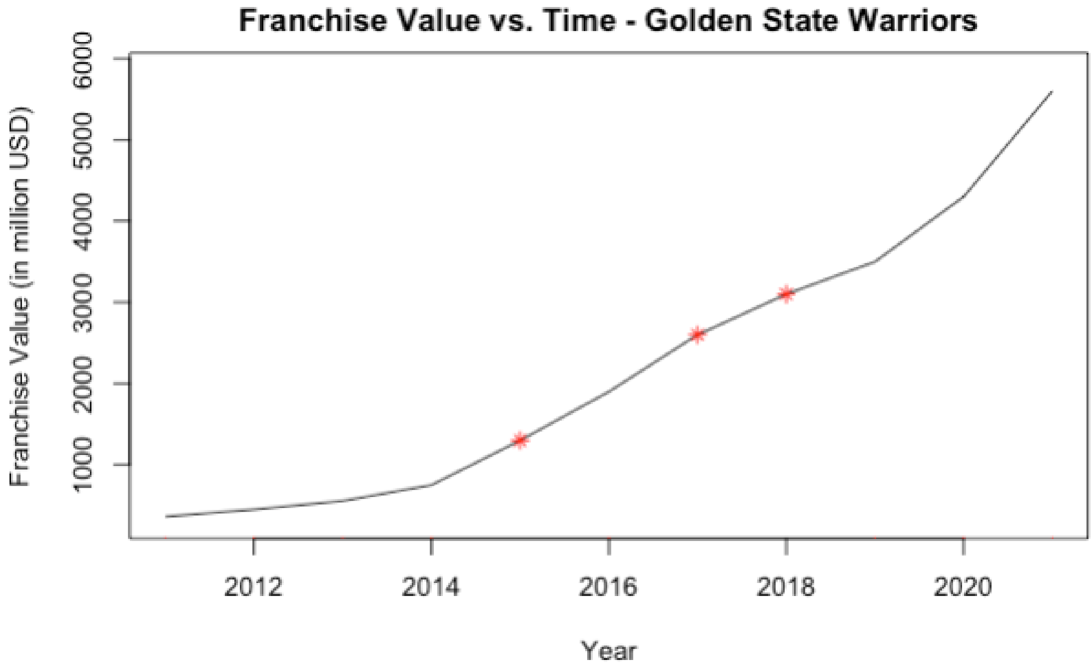


Figure 20: Plot of franchise value versus time for the Golden State Warriors. The red stars indicate years when the Golden State Warriors won a championship.

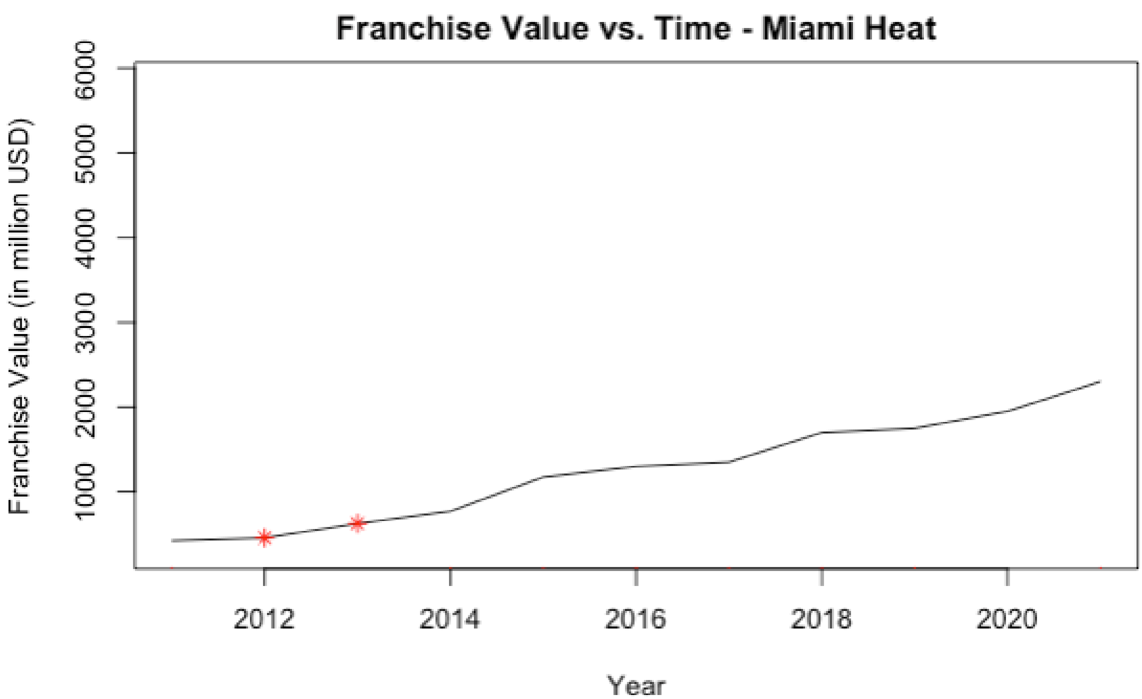


Figure 21: Plot of franchise value versus time for the Miami Heat. The red stars indicate years when the Miami Heat won a championship.

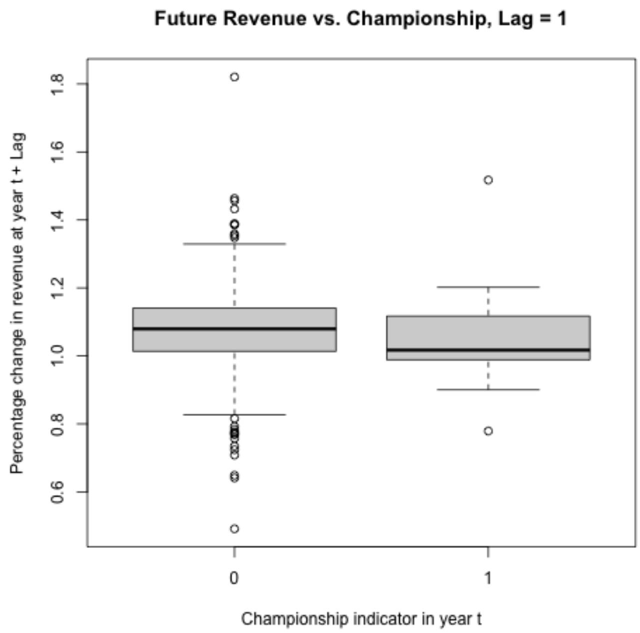


Figure 22: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year $t + \Delta$, for time lag $\Delta = 1$.

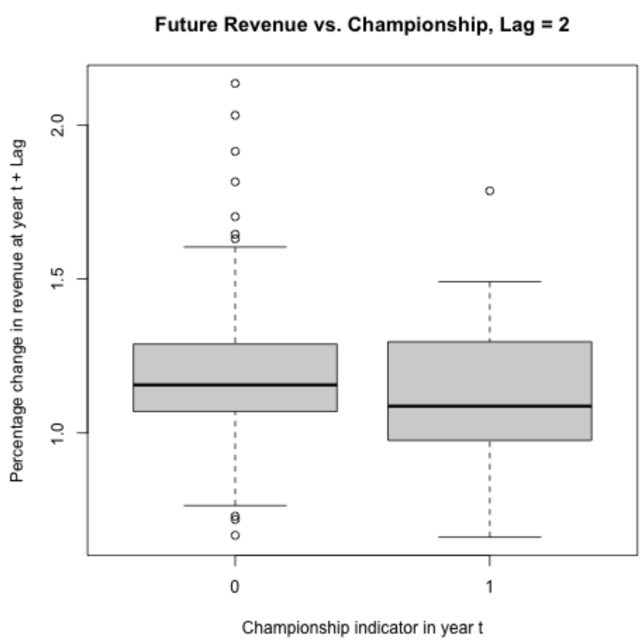


Figure 23: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year $t + \Delta$, for time lag $\Delta = 2$.

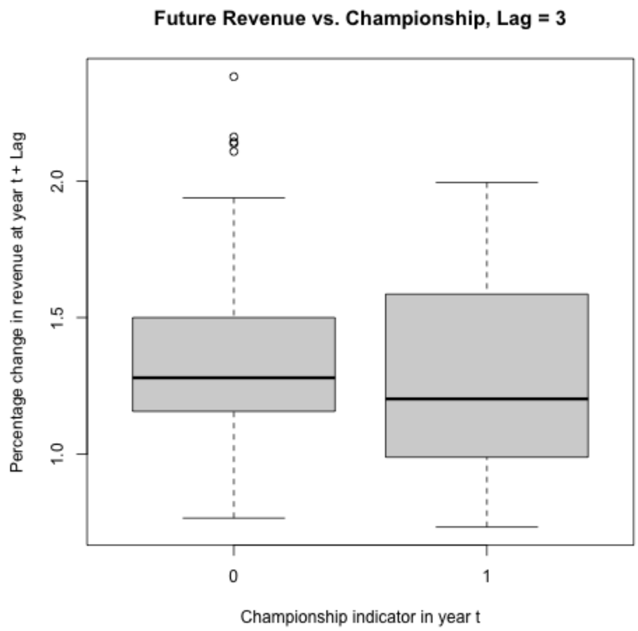


Figure 24: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 3$.

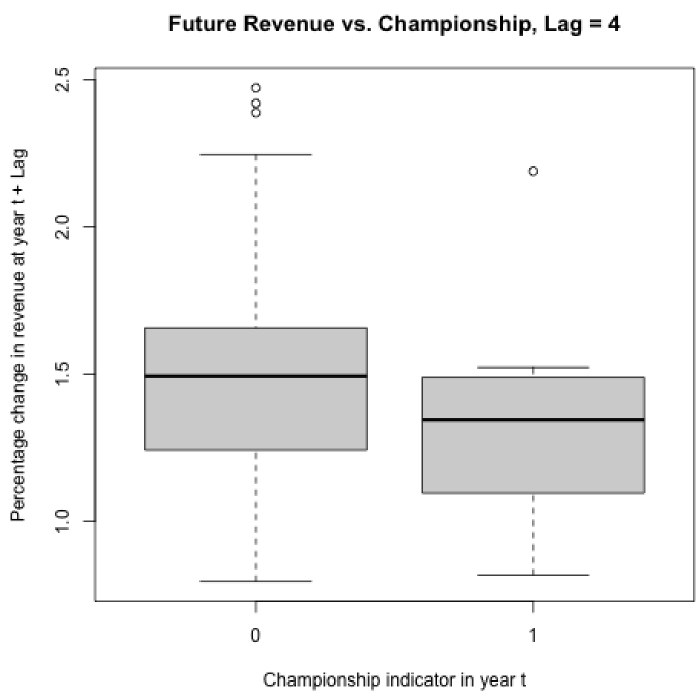


Figure 25: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 4$.

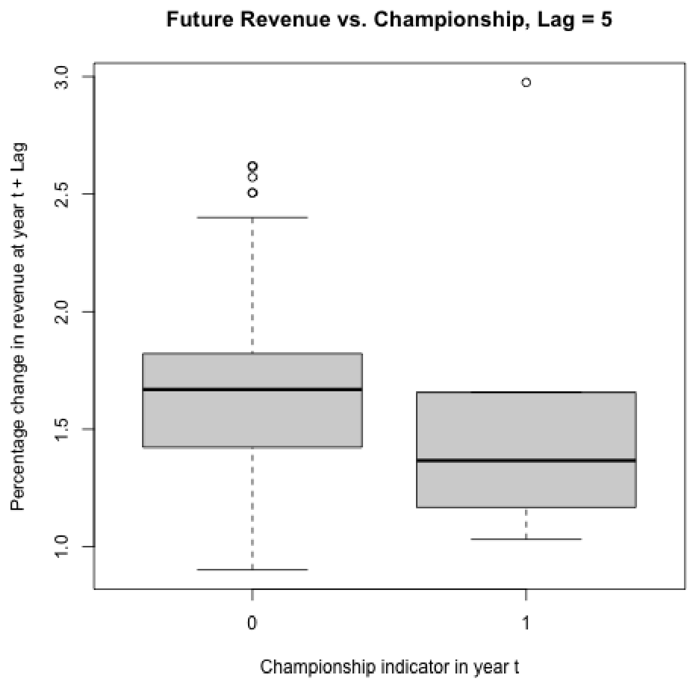


Figure 26: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 5$.

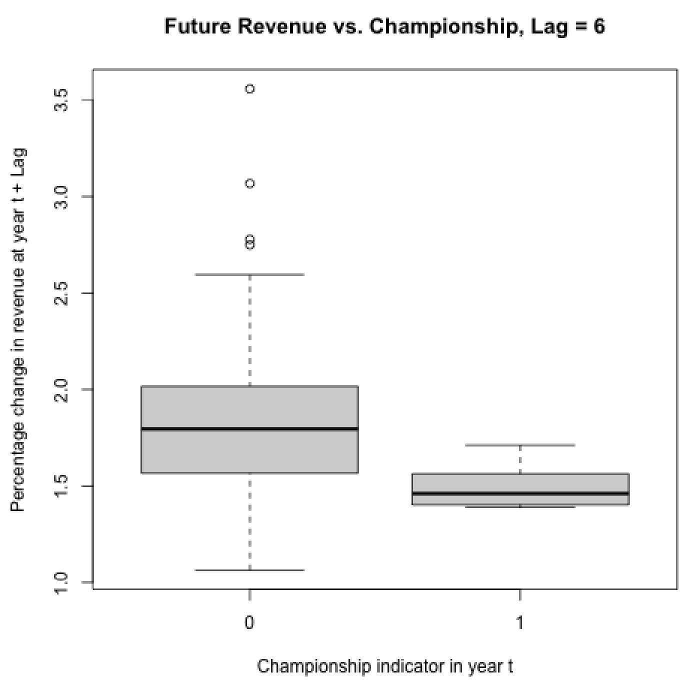


Figure 27: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 6$.

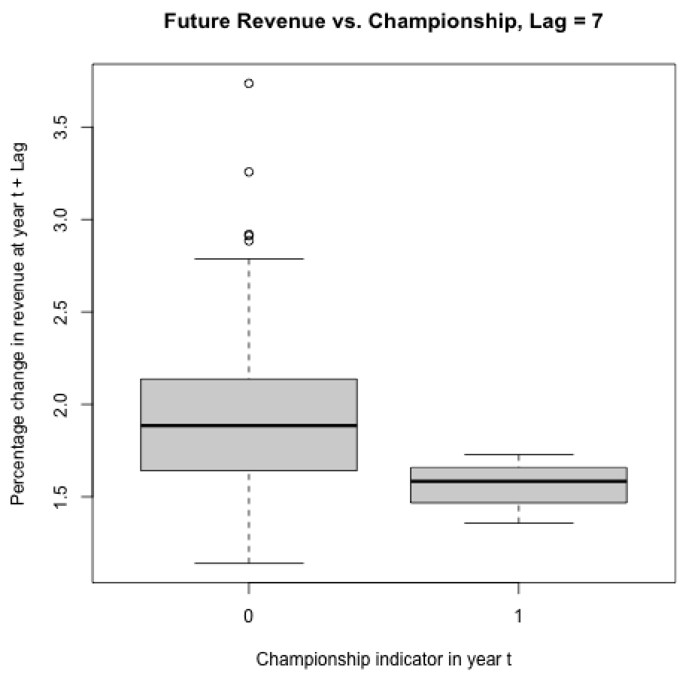


Figure 28: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 7$.

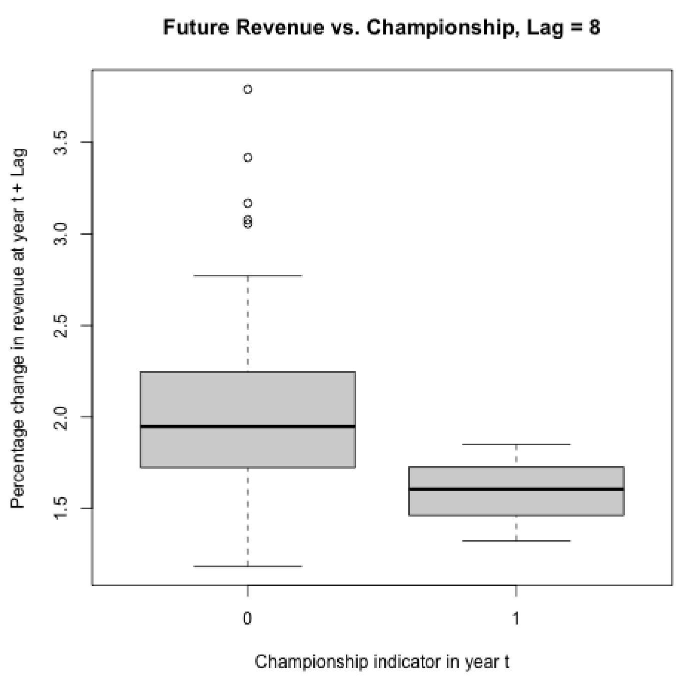


Figure 29: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 8$.

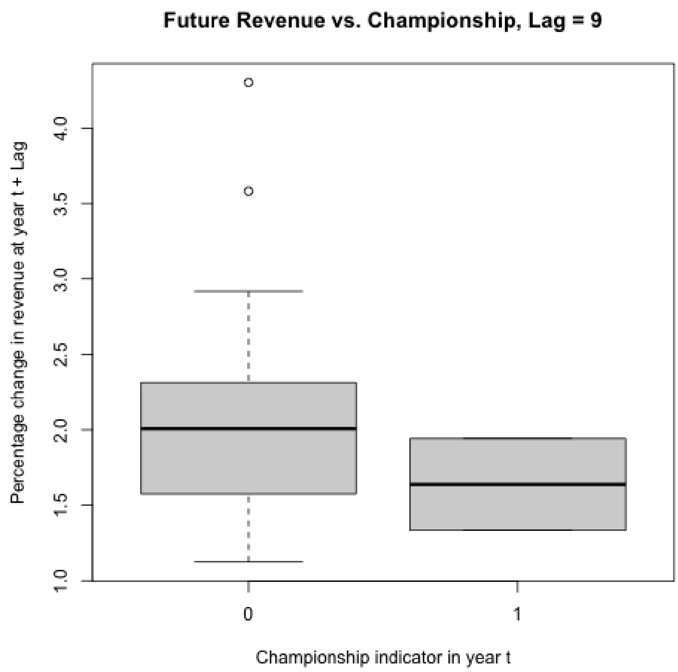


Figure 30: Boxplot comparing percent changes in revenue between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 9$.

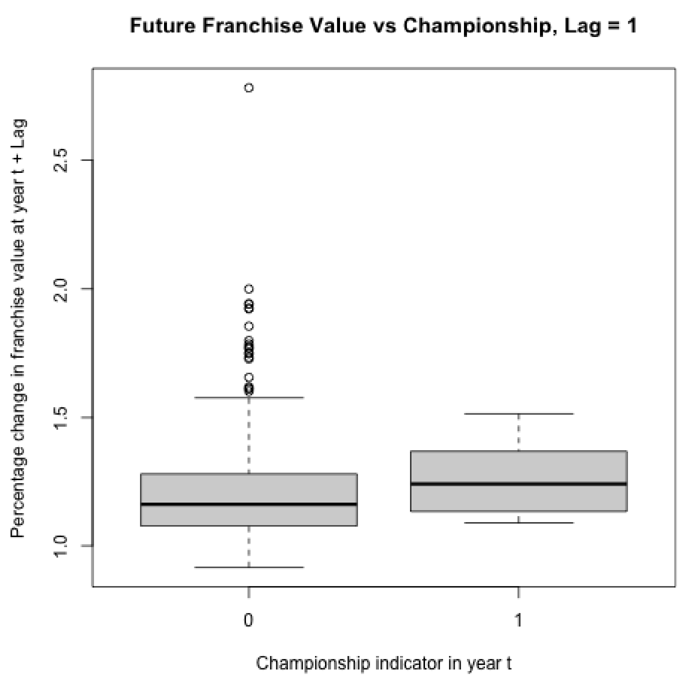


Figure 31: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 1$.

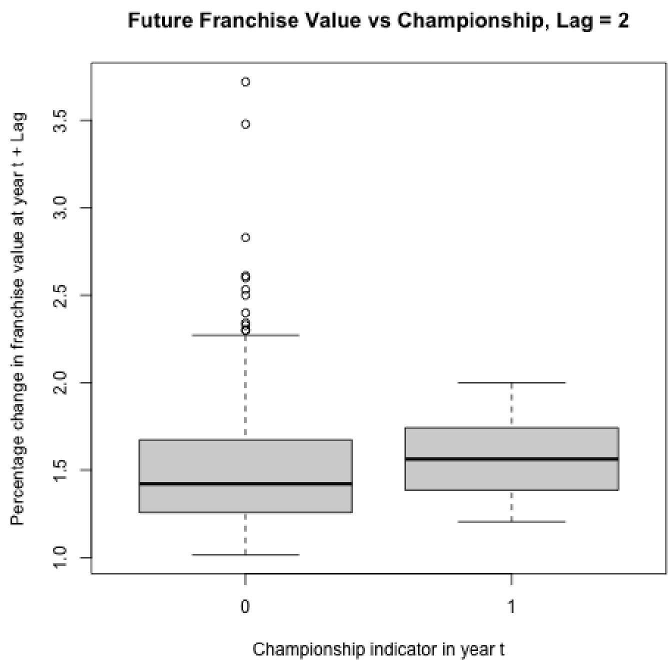


Figure 32: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 2$.

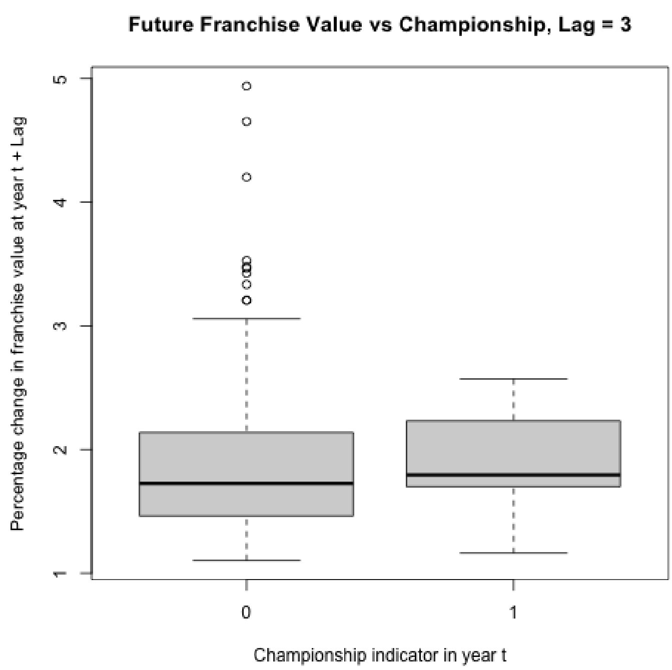


Figure 33: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 3$.

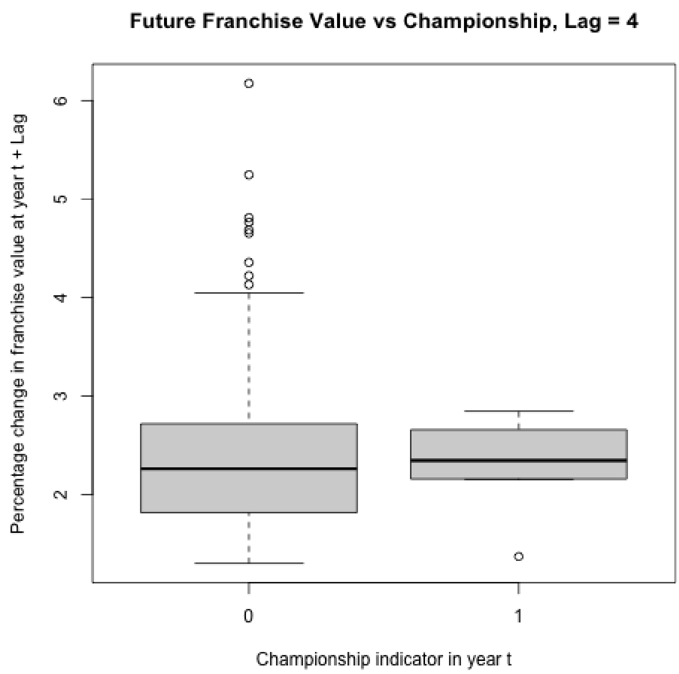


Figure 34: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 4$.

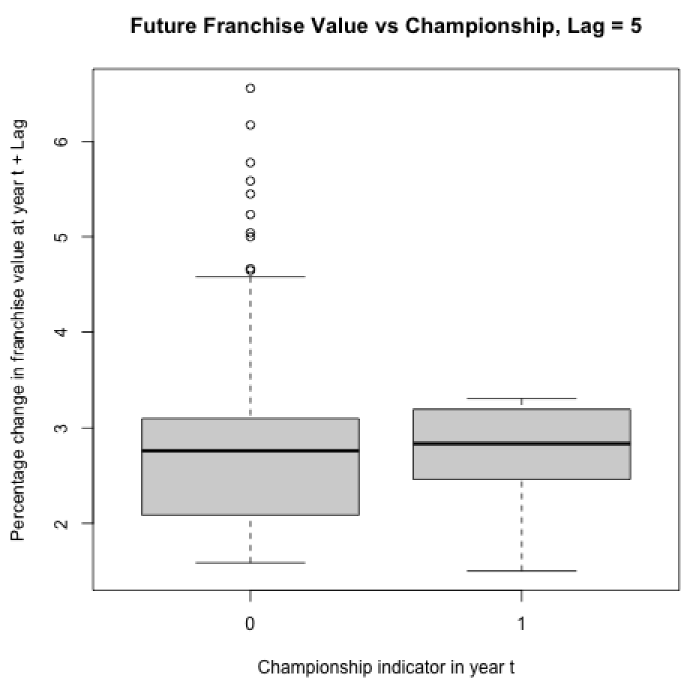


Figure 35: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 5$.

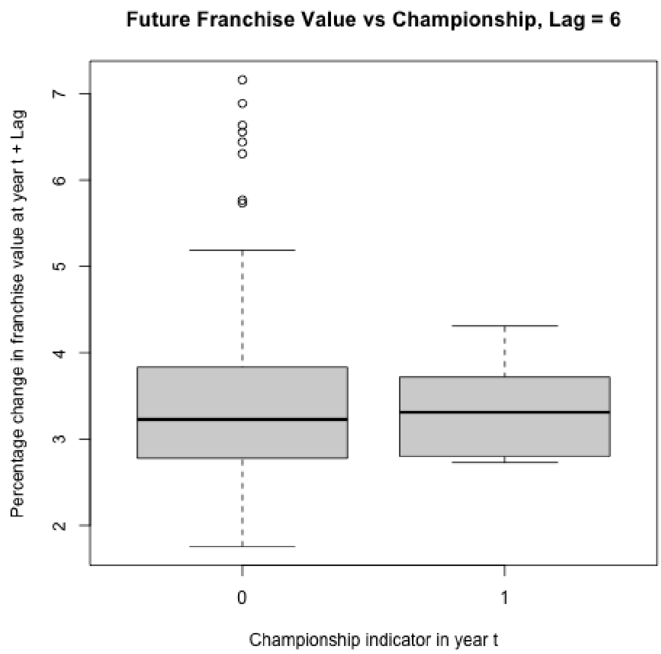


Figure 36: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 6$.

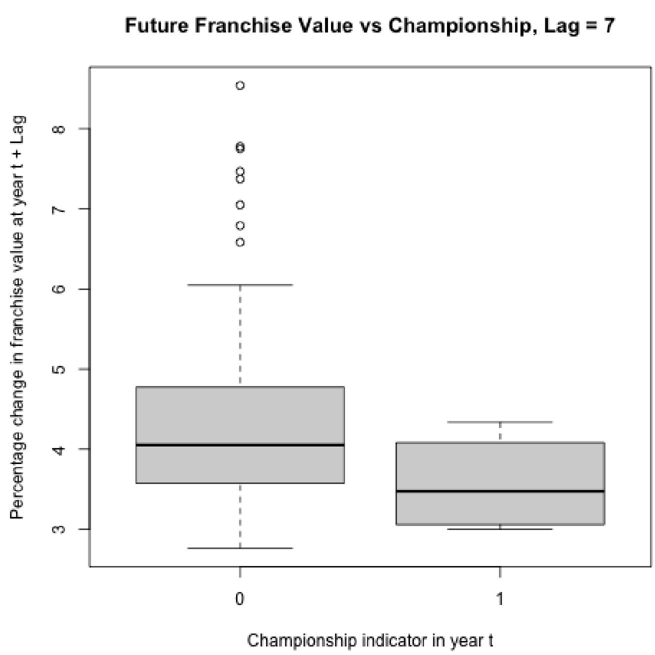


Figure 37: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag $\Delta = 7$.

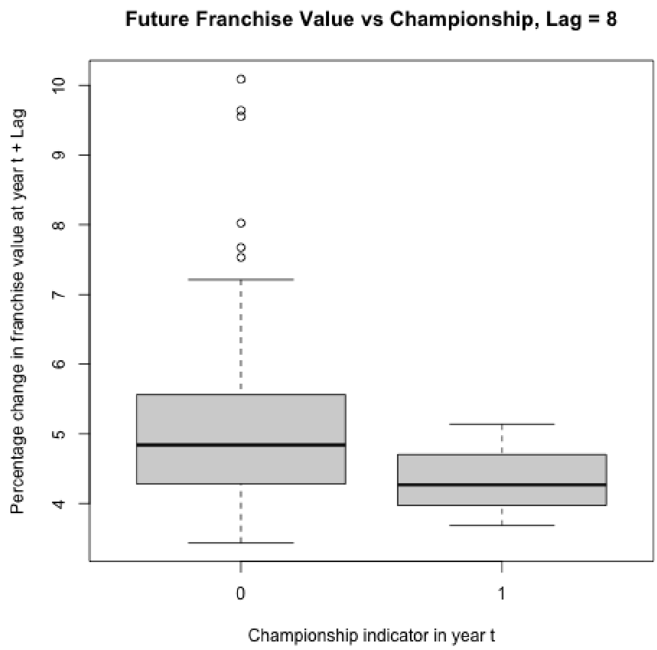


Figure 38: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag 8.

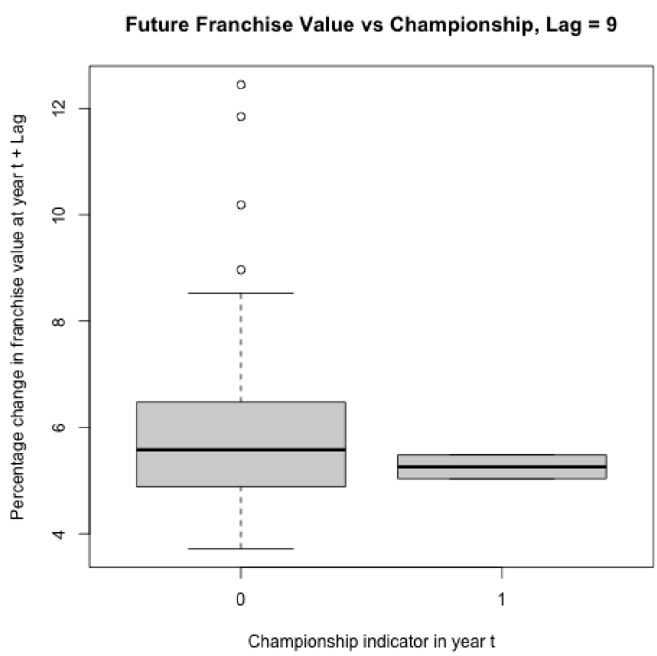


Figure 39: Boxplot comparing percent changes in franchise value between winners and non-winners from year *t* to year *t* + Δ, for time lag 9.

Lag	# of Samples	# of Championship Samples	P-vals, *Revenue*	Reject Null, *Revenue*	P-Vals, *French_Val*	Reject Null, *Franch_Val*
1	300	10	0.581	False	0.179	False
2	270	9	0.551	False	0.220	False
3	240	8	0.576	False	0.492	False
4	210	7	0.741	False	0.621	False
5	180	6	0.576	False	0.662	False
6	150	5	0.998	False	0.549	False
7	120	4	0.996	False	0.956	False
8	90	3	0.953	False	0.895	False
9	60	2	0.796	False	0.961	False

Figure 40: Results from one-sided t-tests between championship and non-championship teams and their future change in revenue and franchise value over some number of lag years. The alternative hypothesis for all the t-tests was that championship teams have significantly larger increases in revenue in the following years compared to non-championship teams.

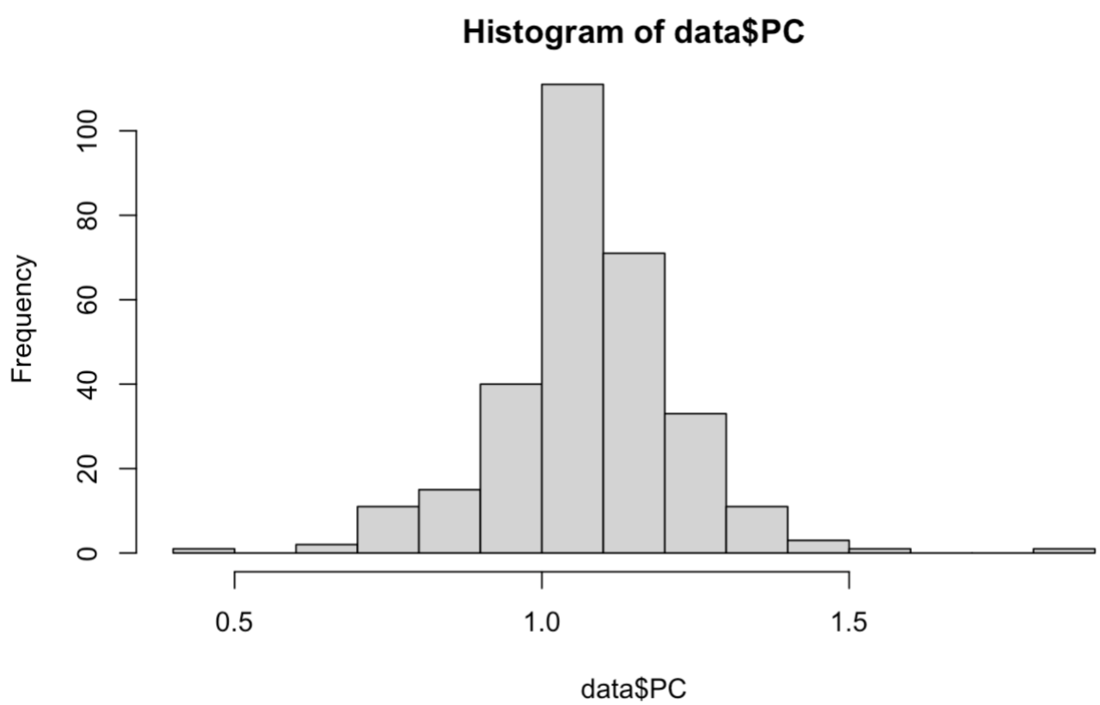


Figure 41: Histogram of response variable *PC*

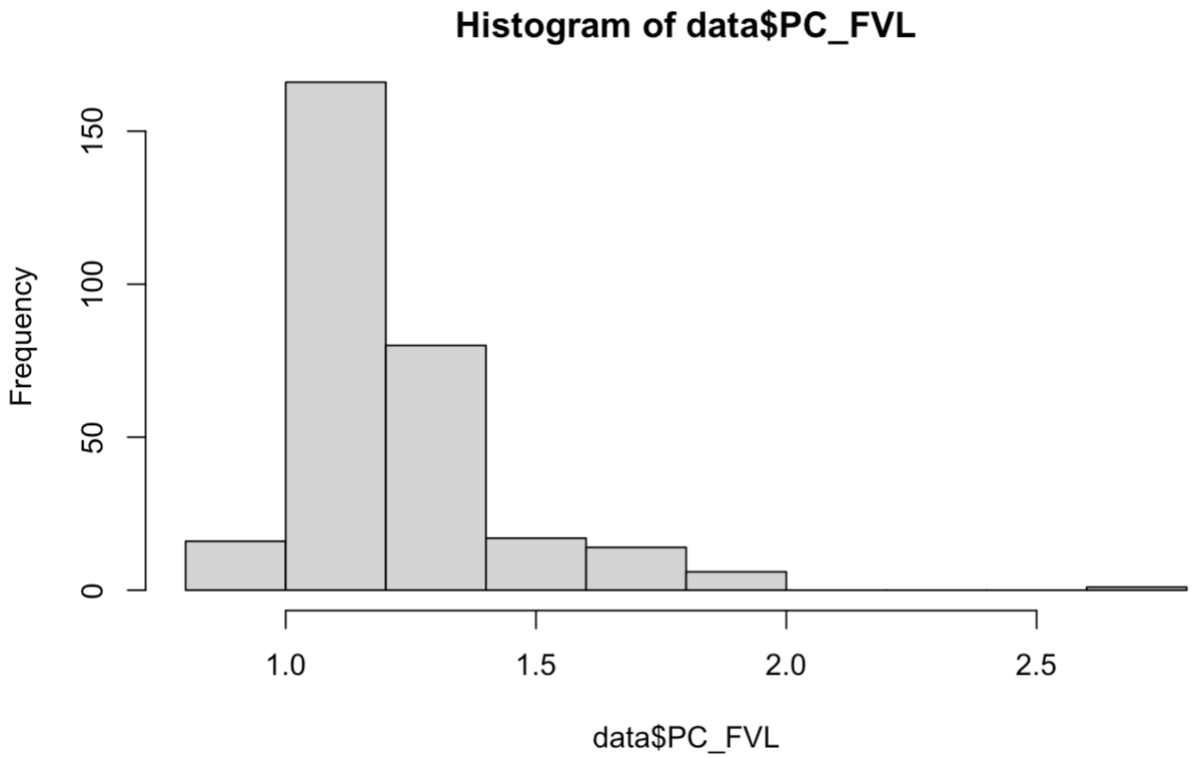


Figure 42: Histogram of response variable *PC_FVL*

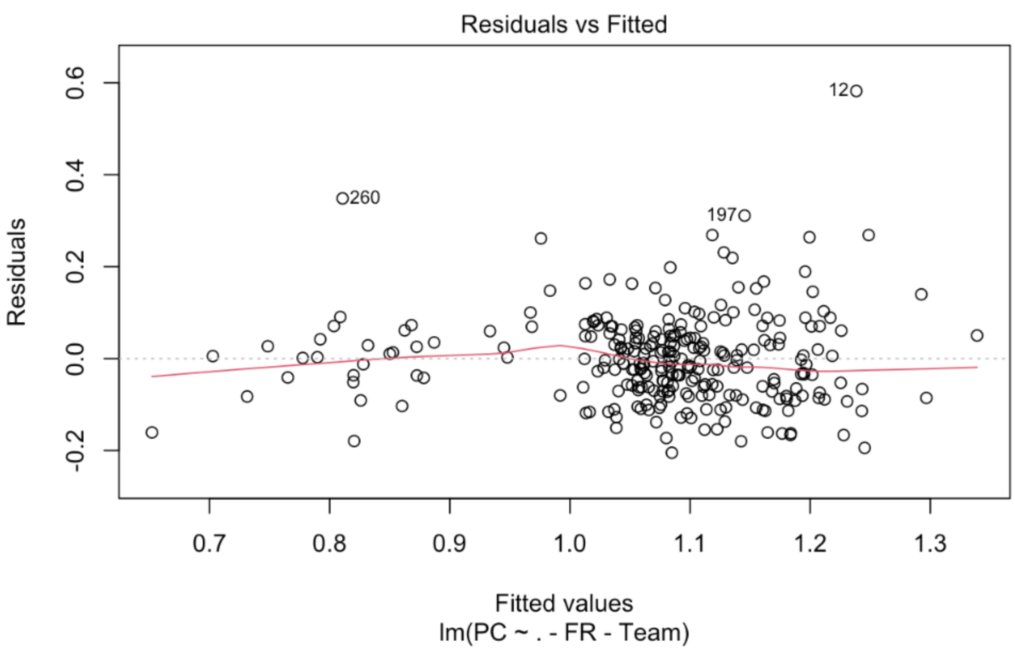


Figure 43: Plot of fitted versus residual values for the *PC* baseline model. Observation 12 corresponds to the 2012 Brooklyn Nets and observation 260 corresponds to the 2020 Utah Jazz.

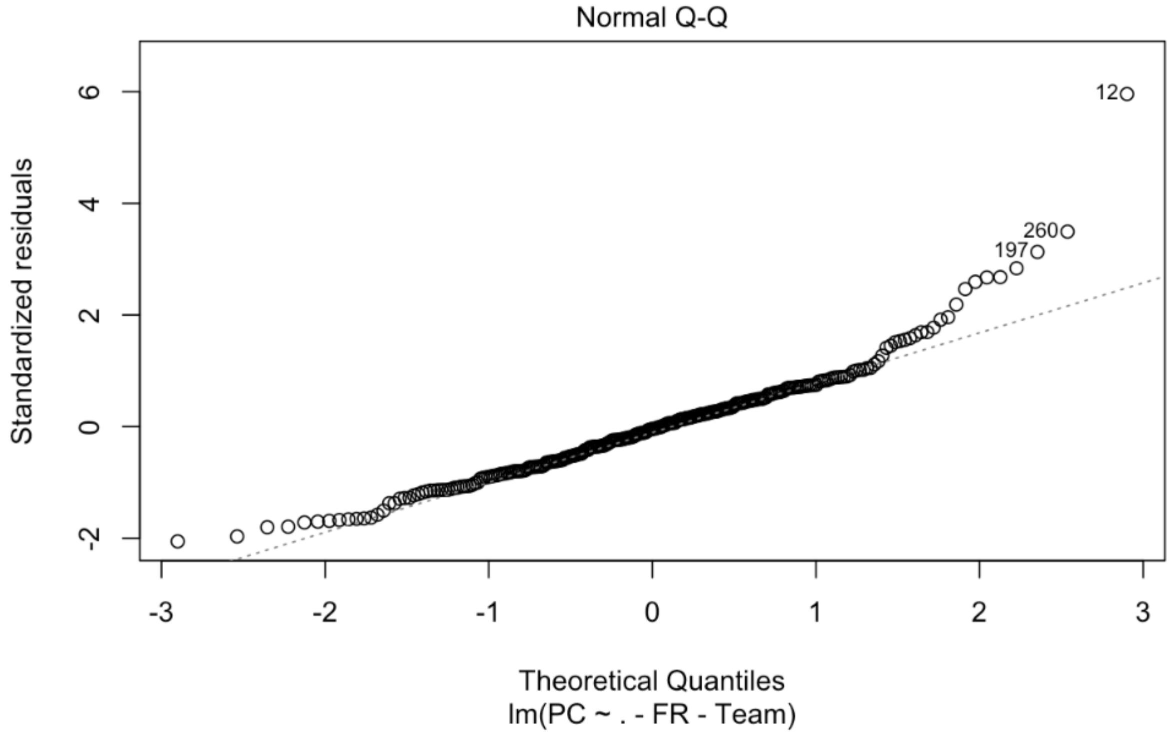


Figure 44: Q-Q Plot for the *PC* baseline model. Observation 12 corresponds to the 2012 Brooklyn Nets and observation 260 corresponds to the 2020 Utah Jazz.

```
Call:
lm(formula = PC ~ . - FR - Team, data = data.train.revenue)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.19242 -0.06895 -0.00790  0.05247  0.32567 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.818e+01  1.738e+01   1.046   0.2967    
Year         -8.656e-03  8.633e-03  -1.003   0.3170    
Revenue      -3.243e-01  6.152e-02  -5.272 2.96e-07 ***
Ticket       -3.855e-02  3.367e-02  -1.145   0.2533    
OI           -6.358e-04  3.022e-04  -2.104   0.0364 *  
Population   -4.589e-03  1.229e-02  -0.373   0.7093    
Income        1.004e-01  5.262e-02   1.907   0.0577 .  
Age          -1.228e-02  4.841e-03  -2.538   0.0118 *  
APG           3.704e-06  4.665e-06   0.794   0.4279    
Playoffs     -2.255e-02  2.024e-02  -1.114   0.2665    
CSF          -2.295e-02  1.976e-02  -1.162   0.2464    
CF           -2.300e-02  3.127e-02  -0.736   0.4627    
Finals        6.802e-02  4.455e-02   1.527   0.1281    
Championship -4.007e-02  4.509e-02  -0.889   0.3751    
WP            2.021e-01  7.962e-02   2.538   0.0118 *  
MPC1         -2.152e-01  6.469e-01  -0.333   0.7396    
MPC2         -1.593e-01  2.421e-01  -0.658   0.5112    
MPC3          6.835e-02  5.074e-01   0.135   0.8930    
Followers     1.614e-02  1.089e-02   1.482   0.1395    
Franch_Val    2.102e-01  4.001e-02   5.254 3.22e-07 ***
COVID        -2.964e-01  6.447e-02  -4.598 6.84e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0934 on 246 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5918,	Adjusted R-squared:  0.5586 
F-statistic: 17.83 on 20 and 246 DF,  p-value: < 2.2e-16

[1] "BIC:"
[1] -407.2737
```

Figure 45: R Summary of the *PC* baseline model and BIC on the training data.

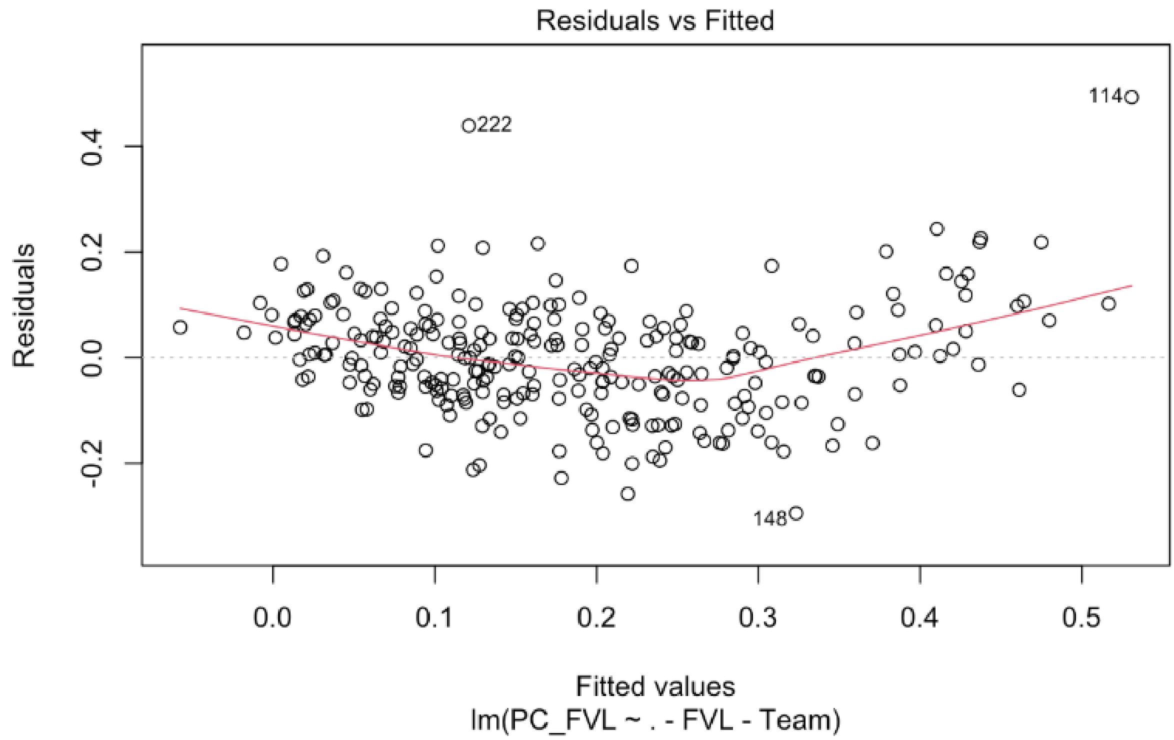


Figure 46: Plot of fitted versus residual values for the *PC_FVL* baseline model. Observation 222 corresponds to the 2012 Sacramento Kings and observation 114 corresponds to the 2014 Los Angeles Clippers.

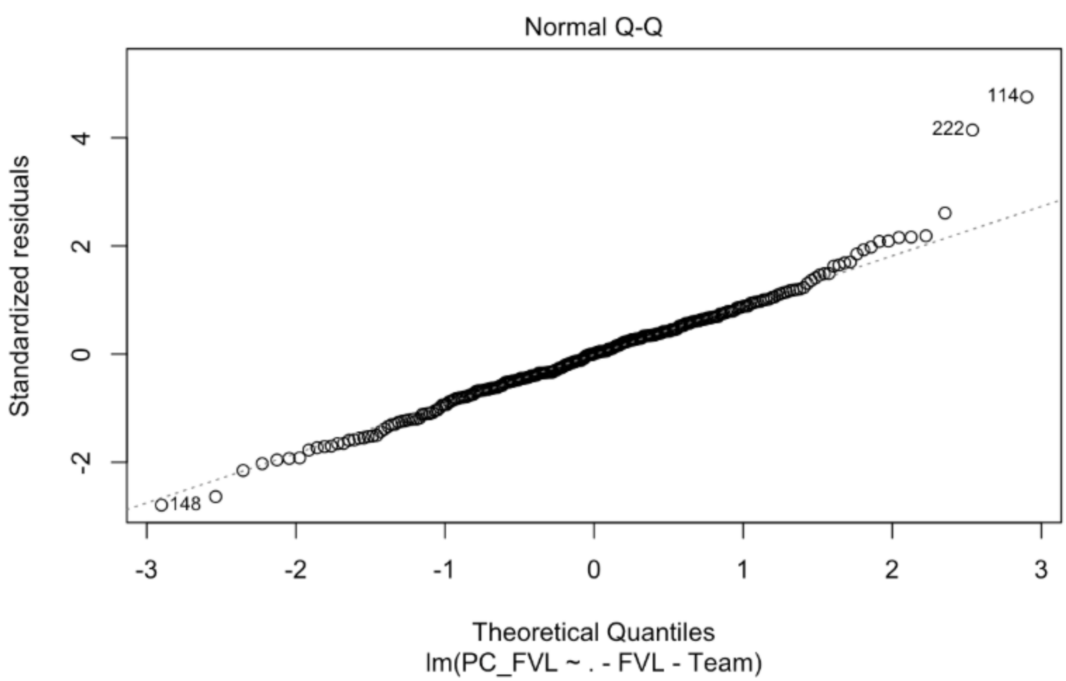


Figure 47: Q-Q plot for the *PC_FVL* baseline model. Observation 222 corresponds to the 2012 Sacramento Kings and observation 114 corresponds to the 2014 Los Angeles Clippers.

```
Call:
lm(formula = PC_FVL ~ . - FVL - Team, data = data.train.franch.val)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.283152 -0.062245  0.002044  0.062825  0.247462 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.580e+01  1.881e+01  -0.840  0.40170    
Year          7.117e-03  9.345e-03   0.762  0.44702    
Revenue       1.077e-01  6.595e-02   1.633  0.10364    
Ticket       -3.657e-02  3.641e-02  -1.004  0.31620    
OI            4.930e-04  3.240e-04   1.522  0.12939    
Population    3.231e-02  1.308e-02   2.470  0.01419 *  
Income        2.239e-01  5.627e-02   3.979 9.13e-05 ***
Age          -5.171e-03  5.184e-03  -0.998  0.31949    
APG           2.221e-06  5.016e-06   0.443  0.65831    
Playoffs      7.186e-03  2.166e-02   0.332  0.74039    
CSF           2.373e-03  2.127e-02   0.112  0.91125    
CF           -2.337e-02  3.358e-02  -0.696  0.48709    
Finals       -4.132e-02  4.773e-02  -0.866  0.38760    
Championship  4.462e-02  4.831e-02   0.924  0.35663    
WP            1.155e-01  8.539e-02   1.353  0.17722    
MPC1         -2.257e+00  6.911e-01  -3.266  0.00125 ** 
MPC2         -7.156e-01  2.595e-01  -2.758  0.00625 ** 
MPC3          3.258e+00  5.434e-01   5.996 7.15e-09 ***
Followers     6.063e-02  1.162e-02   5.218 3.84e-07 ***
Franch_Val   -2.629e-01  4.309e-02  -6.103 4.03e-09 ***
COVID        -4.450e-01  6.909e-02  -6.441 6.20e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1001 on 246 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.584,	Adjusted R-squared:  0.5502 
F-statistic: 17.27 on 20 and 246 DF,  p-value: < 2.2e-16

[1] "BIC:"
[1] -370.3148
```

Figure 48: R Summary of the *PC_FVL* baseline model and BIC on the training data.

```
Call:
lm(formula = PC ~ 1, data = data.train.revenue)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.58131 -0.06520  0.00558  0.06629  0.44472 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.07270    0.00862   124.4   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1411 on 267 degrees of freedom

[1] "BIC:"
[1] -278.8791
```

Figure 49: R Summary of the *PC* intercept-only model and BIC on the training data.

```
Call:
lm(formula = PC_FVL ~ 1, data = data.train.franch.val)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.26895 -0.10075 -0.02885  0.06592  0.51275 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.180398   0.009099   19.82   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.149 on 267 degrees of freedom

[1] "BIC:"
[1] -249.8357
```

Figure 50: R Summary of the *PC_FVL* intercept-only model and BIC on the training data.

```
Call:
lm(formula = PC ~ (. - FR - Team)^2, data = data.train.revenue)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.13062 -0.02163  0.00000  0.01862  0.18216 


Residual standard error: 0.06531 on 89 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9278,	Adjusted R-squared:  0.7842 
F-statistic:  6.46 on 177 and 89 DF,  p-value: < 2.2e-16

[1] "BIC:"
[1] 7.455762
```

Figure 51: R Summary of the *PC* full interaction model and BIC on the training data. The coefficient estimates are not displayed given how many there are.

```
Call:
lm(formula = PC_FVL ~ (. - FVL - Team)^2, data = data.train.franch.val)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.08699 -0.01851  0.00000  0.02071  0.08289 


Residual standard error: 0.05595 on 89 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.953,	Adjusted R-squared:  0.8594 
F-statistic: 10.19 on 177 and 89 DF,  p-value: < 2.2e-16

[1] "BIC:"
[1] -75.17042
```

Figure 52: R Summary of the *PC_FVL* full interaction model and BIC on the training data. The coefficient estimates are not displayed given how many there are.

```
lm(formula = PC ~ COVID, data = data.train.revenue)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.30874 -0.07214 -0.00838  0.04963  0.41543 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.101981   0.007025  156.86   <2e-16 ***
COVID       -0.301858   0.022555  -13.38   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1093 on 266 degrees of freedom
Multiple R-squared:  0.4024,	Adjusted R-squared:  0.4001 
F-statistic: 179.1 on 1 and 266 DF,  p-value: < 2.2e-16

[1] "BIC:"
[1] -411.2568
```

Figure 53: R Summary of the *PC* stepwise model and BIC on the training data.

```
Call:
lm(formula = PC_FVL ~ MPC2 + MPC1 + Income + WP + MPC3 + COVID + 
    MPC2:MPC1 + MPC1:MPC3, data = data.train.franch.val)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.185617 -0.055937 -0.000071  0.051119  0.271530 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.62089    0.35277  -4.595 6.77e-06 ***
MPC2          0.88869    0.41665   2.133 0.033869 *  
MPC1          1.08716    0.62844   1.730 0.084833 .  
Income        0.15363    0.03297   4.659 5.09e-06 ***
WP            0.11142    0.03162   3.524 0.000502 ***
MPC3         -0.60048    0.36529  -1.644 0.101420    
COVID        -0.49521    0.05172  -9.574  < 2e-16 ***
MPC2:MPC1   -72.03218   11.47498  -6.277 1.44e-09 ***
MPC1:MPC3    89.30820    9.99124   8.939  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07872 on 259 degrees of freedom
Multiple R-squared:  0.7291,	Adjusted R-squared:  0.7208 
F-statistic: 87.14 on 8 and 259 DF,  p-value: < 2.2e-16

[1] "BIC:"
[1] -555.1409
```

Figure 54: R Summary of the *PC_FVL* stepwise model and BIC on the training data.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: PC ~ (1 | Year)
   Data: data.train.revenue

     AIC      BIC   logLik deviance df.resid
  -525.1   -514.4    265.6   -531.1      265

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7644 -0.5722 -0.1301  0.4079  4.6154

Random effects:
 Groups   Name        Variance Std.Dev.
 Year     (Intercept) 0.013070 0.11433
 Residual             0.006966 0.08346
Number of obs: 268, groups:  Year, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.07137    0.03651   29.34
[1] "BIC:"
[1] -514.3768
[1] "Conditional R Squared:"
[1] 0.6619418
```

Figure 55: R Summary of the *PC* intercept-only mixed effects model.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: PC_FVL ~ (1 | Year)
   Data: data.train.franch.val

     AIC      BIC   logLik deviance df.resid
  -555.0   -544.3    280.5   -561.0      265

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.3756 -0.6230 -0.0908  0.6515  3.7337

Random effects:
 Groups   Name        Variance Std.Dev.
 Year     (Intercept) 0.016263 0.12753
 Residual             0.006153 0.07844
Number of obs: 268, groups:  Year, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.18184    0.04061   4.478
[1] "BIC:"
[1] -544.2669
[1] "Conditional R Squared:"
[1] 0.7320648
```

Figure 56: R Summary of the *PC_FVL* intercept-only mixed effects model.

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: PC_FVL ~ Income + WP + (1 | Year)
   Data: data.train.franch.val

     AIC      BIC   logLik deviance df.resid
    -588     -570      299     -598      263

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6354 -0.6164 -0.0363  0.6959  3.6412

Random effects:
 Groups   Name        Variance Std.Dev.
 Year     (Intercept) 0.016350 0.12787
 Residual             0.005332 0.07302
Number of obs: 268, groups:  Year, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept) -1.50733    0.32887  -4.583
Income       0.15327    0.03059   5.011
WP           0.10946    0.02933   3.731

Correlation of Fixed Effects:
       (Intr) Income
Income -0.991
WP     -0.027 -0.017
[1] "BIC:"
[1] -570.0007
[1] "Conditional R Squared:"
[1] 0.7678112
```

Figure 57: R Summary of the *PC_FVL* mixed effects model with fixed effects and random intercept.