Final Paper

Playoff Participation and Revenues in Major League Baseball

Spring 2018

Mack Darrow
R. Scott Munro
Bharat Sinha
Sahil Bhatia
Huan Liang

“We pledge our honor that we have not violated the Booth Honor Code during this assignment.”

Table of Contents

Executive Summary

In this paper, we explore the relationship between attendance revenue and playoff participation for American Major League Baseball (MLB) teams. Others have previously explored this relationship, however, their focus was not on playoff participation but on winning more generally. Silver et al. hypothesize that a baseball team's revenue is strongly dependent on three factors - How much the team 'Wins', the 'Market' in which the team is based and the age of the team's 'Stadium'. We build on their work in this paper. For baseline comparison, we recreated their model by calculating team attendance revenue data using the Fan Cost Index and attendance data from Baseball Reference, Stadium Age data from Wikipedia, and the market data (Population and Per Capita Income) from the Federal Reserve Bank.

We improve upon their work by doing **three** things. **First,** we sharpened their basic OLS regression relating revenue to wins, market and stadium. We found additional significant team and year fixed effects that explain the relationship better than their basic OLS regression. While diving deeper into how stadium age affects revenue, our analysis showed that a new stadium increases revenues for only up to seven years, in contrast to the ten years that was previously hypothesized. We also found that in addition to regular season wins, playoff participation in the previous four years also had a significant impact on revenue while World Series participation was found to be insignificant. **Second**, we more thoroughly explored the relationship between revenue and winning, because we realize these don't necessarily have a linearly dependent relationship. There could be causality or reverse causality involved - more wins could cause higher revenues but it's also possible that higher revenues could enable the teams to buy better players, which could cause an increase in wins. To tease out this causal relationship, we explored an Instrument Variable (IV) regression by analyzing 4 different instrumental variables. We explored the strengths and weaknesses of each and determined that the Baseball Reference Strength of Schedule was the most powerful. Using it as an instrument, we ran IV regressions to understand the causal relationship between playoff participation and revenue but the results showed playoff participations to be insignificant. This could be either due to a weak instrument OR because there is truly no causal relationship between going to the postseason and increased revenue in subsequent years, although we'd like to think it's the former. **Third**, we split up the data into teams that just made it to the playoffs and teams that just missed making the playoffs and ran discontinuity regressions on this subset of data. Here too, playoff participation turned out to be insignificant; we hypothesize that this is likely due to the high amount of noise present in a small sample. Overall, we feel that our approach is more thorough in accounting for factors that bias the three primary drivers of attendance revenue than Silver et al. Ideally, MLB front offices could have then used our model as a decision making tool to forecast and/or predict their team revenues in the years after making the playoffs. Had we found statistical significance of playoff participation in driving future revenues, our model could have been combined with player evaluation models to determine if player acquisitions made sense from both a baseball and a business perspective as front offices thought about free agency and the trade deadline.

I. Introduction

July 31st serves as the ultimate poker hand for Major League Baseball executives each year. The trade deadline presents the opportunity for teams to either fold or go all in. The deadline has been described as everything from “a little hectic” by former Minnesota Twins GM Terry Ryan to “a mindf**k” by current Chicago Cubs President of baseball operations Theo Epstein [16], [17]. As front office executives decide whether or not to put all their chips on the table, they have to reliably estimate their chances of making the playoffs, the possible damage inflicted upon a farm system by trading away top prospects, and the urgency of the fanbase and the owner in order to determine how aggressive to be. These challenges are magnified when teams consider taking on a “rental” player on an expiring contract. How much additional payroll should a team reasonably take on to compete when it is unlikely that player will be retained? We believe that teams are short-sighted in exploring these questions, and too often only evaluate the baseball side of the equation without understanding the lagged effects of postseason participation on the business side of the equation. In this paper, we seek to study the impact of playoff qualification on future attendance revenue in order to determine if front offices can increase their risk tolerance when it comes to in-season player acquisitions.

We will begin by examining past studies of attendance revenue drivers in Major League Baseball before conducting OLS regressions and Instrumental Variable Regressions in an effort to better understand reverse causality challenges. Additionally, we will run discontinuity regressions based around teams that narrowly make and miss the postseason to test the significance of playoff qualification in Major League Baseball. This paper may not change the emotional state of general managers around the league on July 31st of each year, but the hope is that front offices will be able to use these results to refine their business calculations as they evaluate personnel decisions and contemplate their place in the hunt for October glory.

This paper seeks to explore the relationship between winning and team revenues, specifically from fan attendance. Previous papers have studied this topic and have explored factors predictive of a team's financial success. Across previous studies on revenue generation, there is consensus that three major factors have predictive power of team revenues: 'Winning', 'Market', and 'Stadium'. 'Winning' is measured by winning percentage in the season. 'Market' incorporates city population and per capita income and 'Stadium' accounts for the age of a team's stadium.

Winning is an ambiguous term and can mean different things for different teams. For the purposes of this paper, we specifically define winning as 'making it to the playoffs'. Data shows that winning teams consistently generate higher revenues than losing teams. Winning is also something that the teams can control year over year, whereas stadium and market are relatively fixed variables and change steadily over prolonged periods of time. At a base level, the desire to win is what motivates players, coaches, fans, and owners. It motivates athletes and coaches to work tirelessly to improve their individual capabilities. Spectators stay engaged and attend games because they want to see a winning product. In Major League Baseball (MLB) 30 teams compete each year for the ultimate 'Win' - the Commissioner's Trophy awarded to the World Series Champions. Realistically though, not all teams aim to win the World Series; the definition of winning can vary according to where teams are in their competitive cycle. Winning for a historically successful organization such as the New York Yankees would mean winning the World Series, but winning for most other franchises may mean anything from maximizing wins to competing for wild card playoff spots and division titles. Given that qualifying for the postseason sufficiently narrows the field of competitors and is a prerequisite for winning the World Series, playoff participation is a viable definition of 'Winning' as it pertains to Major League Baseball. Making the playoffs is a common 'winning' goal for all teams.

A playoff appearance generates a healthy revenue stream for several years. If a team reaches the playoffs, it can count on an additional revenue stream over the next 4 - 5 five years [3]. Additional revenue when a team makes the playoffs comes from the following sources: Increased gate receipts as the team plays more games; additional concessions sales; premium on ticket prices for playoff games; additional and premium on Luxury suites and club seats;

additional merchandise sales; and premiums charged by teams on local broadcast rights. There is also a negative effect on Revenue Sharing, because if a team goes into the playoffs, it makes more revenue, but it has to share this revenue with other teams that don't make it into the playoffs [13]. Teams that make the playoffs in successive seasons don't see as much of a revenue jump as in the first year of making the playoffs. When measured against the value of the 1st playoff appearance, there is an approximately 75% effect for the 2nd consecutive playoff appearance, 50% for the 3rd, and about 30% for the 4th and any consecutive playoff appearances thereafter [3]. This makes sense because the 'novelty' factor of winning or making the playoffs reduces as fans would expect a strong team to make it into the playoffs each season and hence will not be willing to pay ever increasing premiums on playoff tickets and other amenities.

The other two factors are 'Market' and 'Stadium'. Market is defined as the size of the city in which the team is based and is subdivided into the population of the Metropolitan Statistical Area (MSA) of the relevant city and the Per Capita Income of the city. As an example, the Philadelphia Phillies and Cleveland Indians play in similar sized cities, but since the Per Capita Income of Philadelphia is higher than that of Cleveland, the Phillies generate higher revenues than the Indians. In regards to the 'Stadium' variable, revenues of teams increase significantly when they build new stadiums. This is due to the newness effect as people want to experience enhanced stadium features and watch games in the new stadium. This is called the 'Honeymoon effect' and , as suggested by Silver et. al, typically lasts for 10 years. This effect partly explains why teams keep building new stadiums even when the useful life of a current stadium may still be at least a decade long. In particular, we are interested in one measure of on-field performance, which is reaching the playoffs. As part of our study, however, we will also explore the impact of other measures of on-field success, such as winning the World Series or improved regular season performance without necessarily reaching the playoffs.

II. Objectives

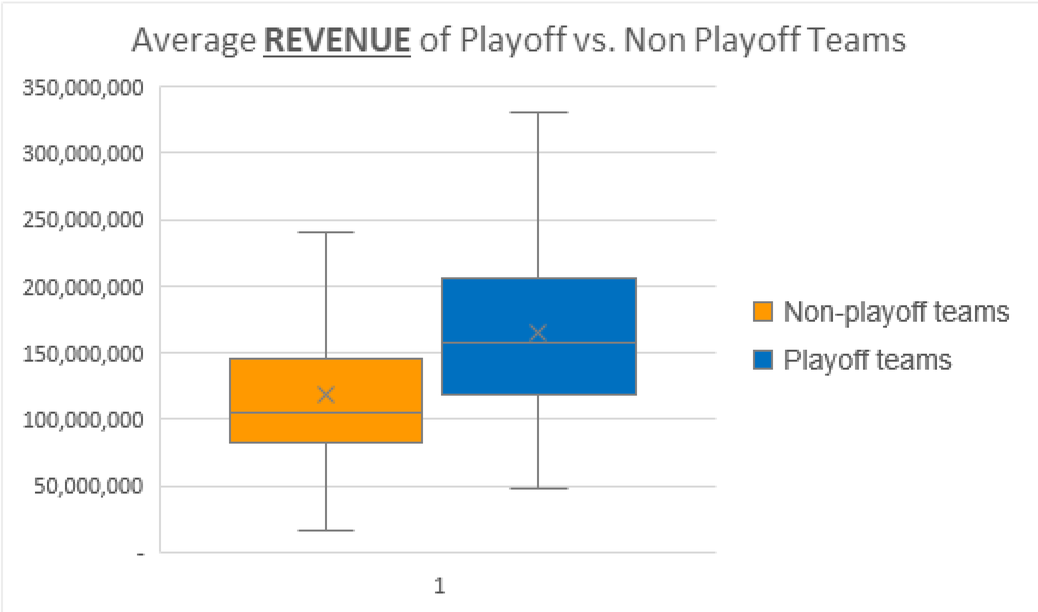


Figure 1a: (Orange) Average of revenues of teams that did not make the playoffs from 1998-2016. (Blue) Average of revenues of teams that made the playoffs from 1998-2016.

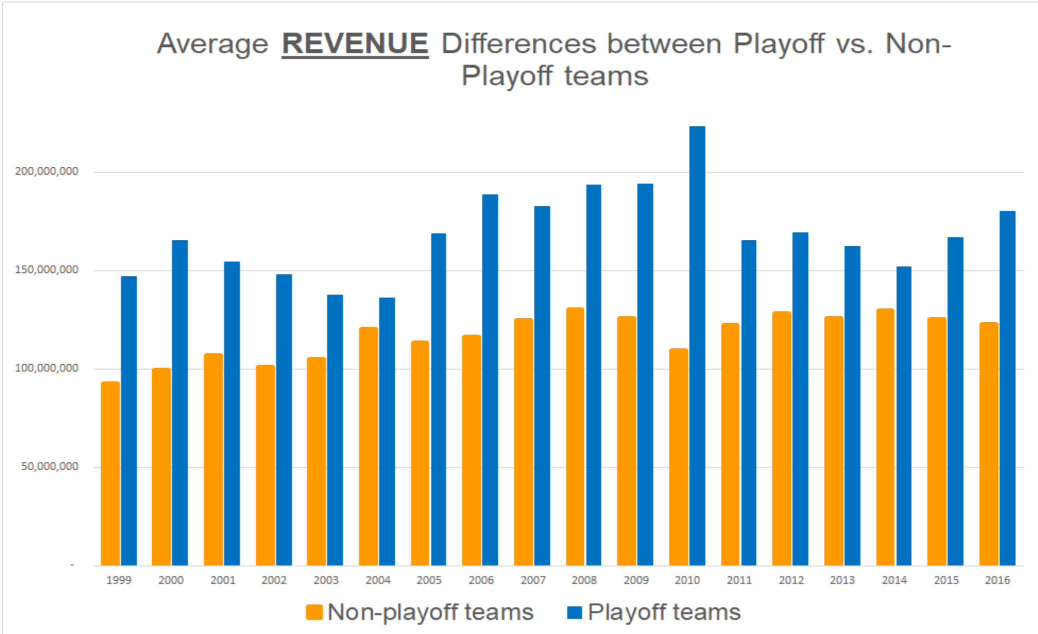


Figure 1b: Year by year comparison of average revenue of teams that made it to the playoffs vs. teams that did not make it to the playoffs.

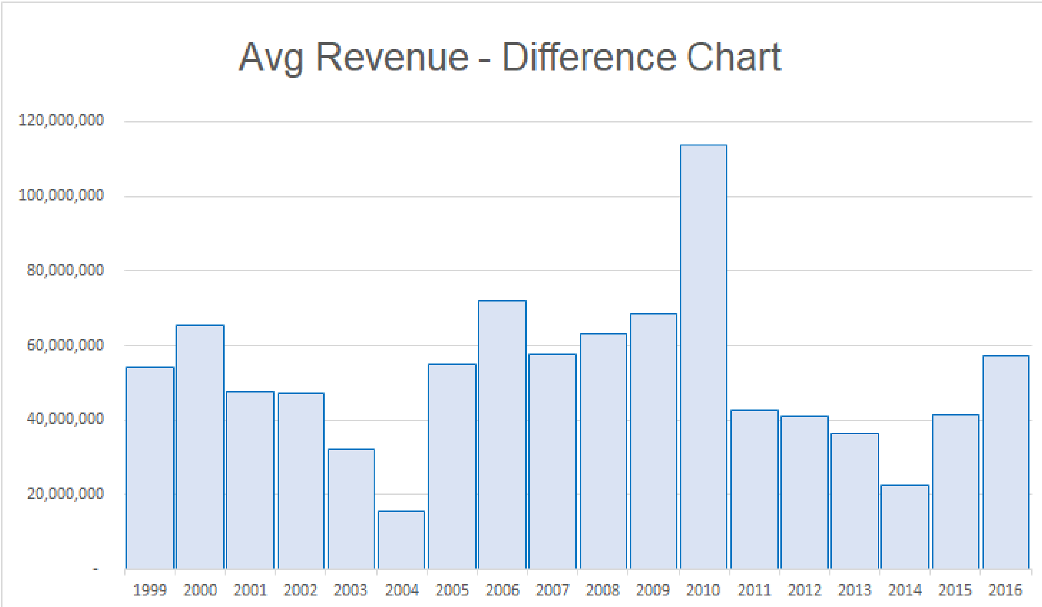


Figure 1c: Year by year differences in average revenue of teams that made it to the playoffs vs. teams that did not make it to the playoffs.

Figures 1a show that there average revenues of teams that make it to the playoffs are higher than revenues of teams who don't make it to the playoffs. When we look at this difference year over year in Figure 1b, we see that this holds true for every year of our dataset. In fact, Figure 1c shows that this number ranges from $22MM to $113MM, with an average of $51MM. Therefore it is important for teams to understand the financial impact of making it to the playoffs. Our motivation for this analysis is to provide a framework that aids the decision making process of General Managers of teams as they are faced with the predicament of determining the merits of additional spending for better players, with the thought being that the better players will buy them more wins or a place in the Playoffs. They can use our model to help them to see how much additional revenue they can get if they spend money buying better players thinking they will get more wins, and evaluating whether the incremental gains of personnel decisions provide sufficient returns at the organizational level. At the trade deadline each year, teams must decide whether they are positioned to compete for the remainder of the season and challenge for a coveted playoff berth. In evaluating whether a roster should be retooled or rebuilt, organizations should consider the revenue effects of qualifying for the postseason in addition to evaluating player attributes and team needs. Under what circumstances should a team consider trading away prized prospects to acquire a pending free agent? At what point do the concerns of a frustrated fanbase need to be addressed with a conscious effort to win? Does momentum generated by the last playoff berth provide a window for teams to develop rookie players before identifying and entering into the next window of contention? By understanding the relationship between playoff participation and revenue in subsequent seasons, GMs and owners can better value decisions by accounting for the net impact of roster alterations.

GMs can use our model to better predict revenues than the Silver et al base model. We consider the following as our Base Model and its equation is as follows:

$$Revenue = A + B\ Wins + C\ Population + D\ Income + E\ Stadium\ Age$$

The base model serves as starting point, uses a honeymoon effect of 10 years and uses regular seasons wins. We recreated the base model using information from Nate Silver's book; our regression results are shown in Appendix A. Our overall objective was to develop a model that performs better than this base model. We try to do this by including additional explanatory variables in an OLS regression - Playoffs, World series, Lag Year Wins, Team Fixed effects and Year Fixed effects.

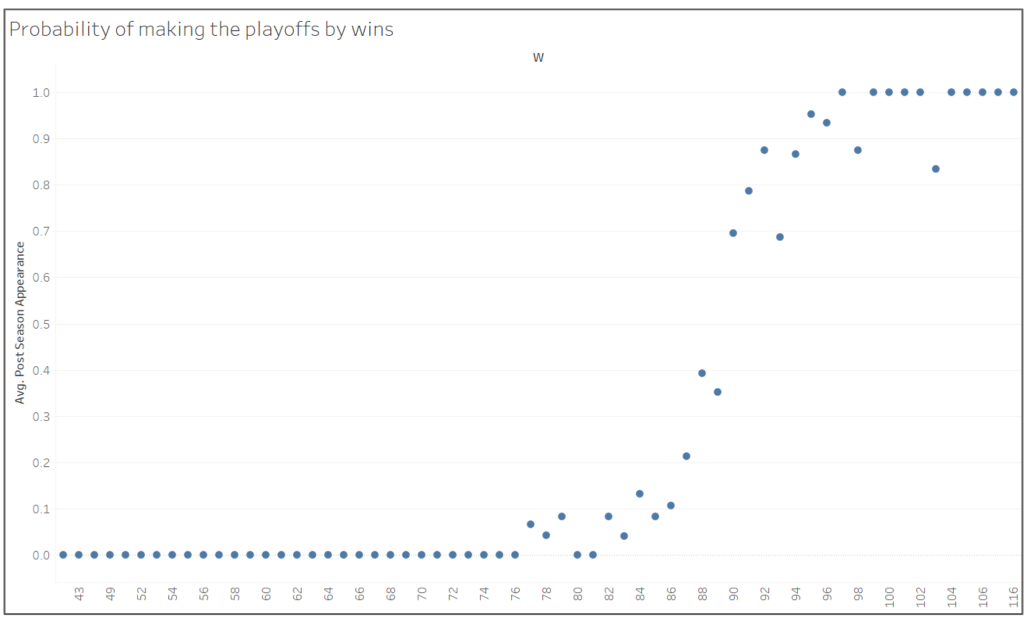


Figure 2: Probability of Playoff Participation and Marginal Value of Additional Wins

We also explore Instrument Variables and Discontinuity regressions to better understand the relationship between revenue and winning. As Figure 2 above shows, teams need to win somewhere in the range of 80-95 games to make it to the playoffs. We look at teams that just made and just missed making it to the playoffs. As an example, in 2008 both the New York Mets and New York Yankees won 89 games in the regular season. However, the Yankees made the playoffs while the Mets did not. By looking specifically at this subset of data, we exploit the inherent randomness in the data to tease out the causal / reverse causal relationship between winning and revenue

III. Economic / Econometric Model

We identified a number of potential models that could help us to tease out the relationship between going to the playoffs and the subsequent impact on revenue. First, we will want to run a basic OLS regression with appropriate controls, in a similar manner to previous authors [1,3]. This approach will prove useful in helping us understand high level relationships within the data. Next, we will use an instrumental variable approach to attempt to tease out the relationship between regular season attendance and revenue. Finally, we will attempt to take advantage of the near randomness between teams that almost make it to the postseason, and teams that do make it to the postseason to see if we can identify statistical differences in the revenue of those teams.

Standard OLS

For our standard OLS regression, we will look to the previous literature in order to craft the appropriate controls, namely: winning controls (both current year and lagged), market controls (per capita income, population, and presence of multiple teams in the market), and stadium controls (age of stadium to capture honeymoon effect).

Thus our model for this simple regression will take the following form:

$$LocalRev_{ij} = \alpha + \beta_{1ij} * Winning_{ij} + \beta_{2ij} * Market_{ij} + \beta_{3ij} * Stadium_{ij} + Team\,FE + Year\,FE + \epsilon \quad (1)$$

In the above model, ‘Winning’, ‘Market’, and ‘Stadium’ represent vectors of controls for team *i* in year *j*. Additionally, we will include team and year fixed effects to allow for differences in years and teams that we have not captured in our other controls. For example, fans in certain cities may be loyal to their teams no matter how badly they play, as evidenced by Chicago's largely unrequited love affair with the Cubs. To capture these differences between franchises, we feel it is appropriate to include team fixed effects. Similarly for year fixed effects, it is very possible that there are changes year over year that are not adequately captured in our controls, and thus year fixed effects will be appropriate in our model. As we discussed, this method has a number of shortcomings, namely that there is likely a correlation between winning and the uncaptured variation in the error term that will systematically bias the coefficients to our winning terms; however, our hope is that we have included enough controls to help alleviate some of this bias. For this reason, we will also be using an instrumental variable method to appropriately adjust for this bias.

Instrumental Variable - Strength of Schedule

Often times in econometric analysis, true randomized controlled trials are impossible. Researchers in this case are forced to rely on observational data alone. In situations where causal inference is the goal, these types of data can be especially challenging as the unobserved factors will no doubt confound the true relationship between the variables in question [6]. Instrumental variables are one method that economists use to get around this issue [7]. While there are a number of potential instruments that would be correlated with winning, and not revenue, the instrument we have decided to focus on is some form of strength of schedule. We looked at four different candidate instruments related to strength of schedule: ESPN's strength of schedule metric, a modified version of ESPN's relative power index (RPI), Baseball Reference's strength of schedule, and an Opponent Win Percentage metric that we calculated using the pythagorean theory of wins.

The instruments themselves are calculated in quite different manners. The ESPN strength of schedule metric is a straightforward combination of opponents' records, and opponents' opponents' records. For ESPN's RPI, because it is initially calculated by combining 25% of a teams' own record, 50% their opponents' record, and 25% their opponents' opponents' record, we modified it by removing the 25% of a teams' own record. Baseball References' strength of schedule metric is the number of runs per game that a team's opponents are better or worse than the average. Finally, the Opponent Win Percentage metric we created is calculated using the pythagorean formula of wins, where the runs scored and runs allowed do not include games in which the team in question played the opponents:

$$Opponent\ Win\ Pct = \frac{RS^{1.82}}{RS^{1.82} + RA^{1.82}}$$

One way to think about whether or not the instruments in question will be good instruments is to see whether or not there is a relationship between the instrument and the probability that a given team will make it to the postseason. To look at this, we broke the instruments into quintiles and looked at the number of teams that made it to the postseason within a given quintile, relative to the total number of teams within that quintile.

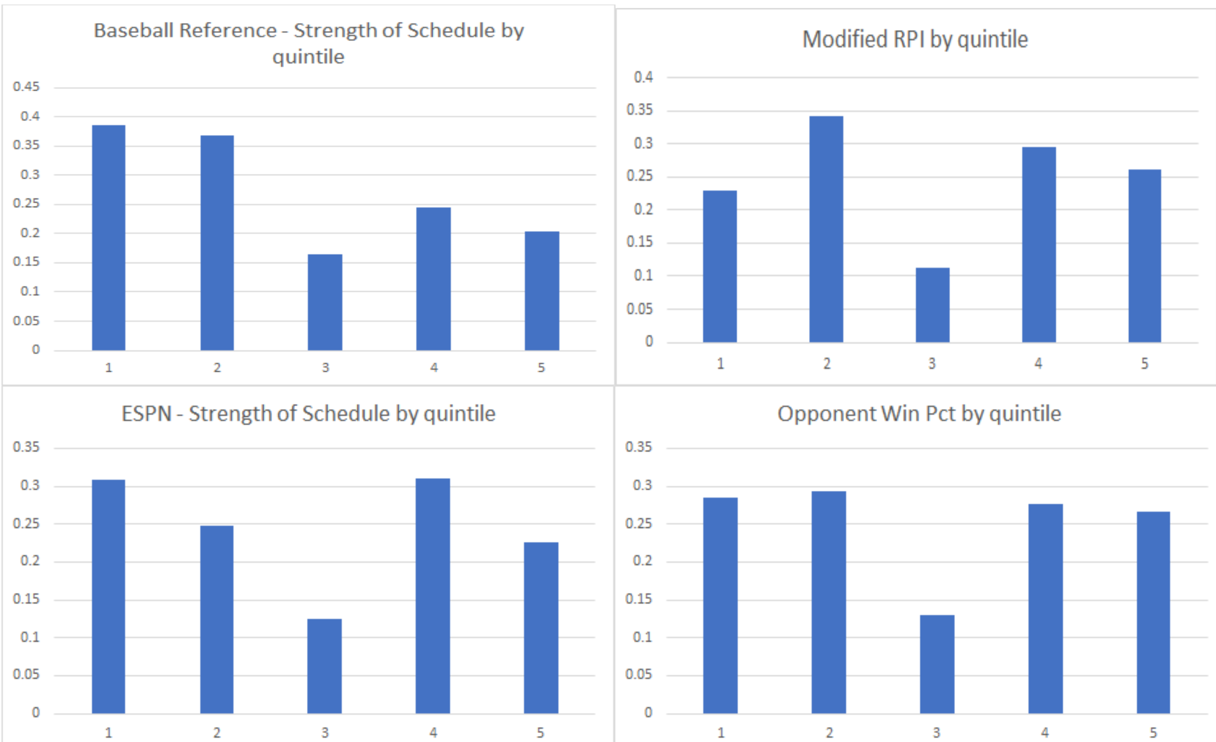


Figure 3: Probability of making it to the playoffs by quintile for different Instruments

Looking at the four instruments in this way does raise some concerns, as there is no clear relationship between the probability that a team makes the postseason and the instruments we have chosen. In fact, there seems to be almost a u-shaped relationship in many of the cases. In other words, the "hardest" and "easiest" quintiles have higher probabilities of making the postseason, whereas the "middle" quintiles have the lowest probability of making the postseason. One further way of looking at how appropriate these various instruments are is to run a series of regressions to see how well each instrument explains the postseason appearances in a given year.

	(1)	(2)	(3)	(4)
ESPN SoS	-3.250 (2.899)			
BR SoS		**-0.253*** (0.093)**		
Opp Win Pct			-1.796 (4.607)	
Modified RSI				0.031 (3.490)
Constant	1.897 (1.449)	**0.274*** (0.020)**	1.173 (2.303)	0.261 (1.310)
Observations	429	509	509	429
R2	.003	.014	.003	.000
Adjusted R2	.001	.012	-.002	-.002
F-Statistic	1.257	7.369	0.152	.001
*p<0.1; **p<0.05; ***p<0.01				

Table 1: Comparison of different instruments

Given the results of the regressions above, we have concerns about all of the instruments with the exception of the Baseball References strength of schedule metric, which has an F-Statistic of 7.369. In an ideal scenario, this instrument would be an even better predictor of post season appearances.

For the instrumental variables that we determine to be statistically significant, we will run a series of regressions where the first stage of the regression, regresses the instrument, z, on whether or not the team made it into the playoffs which is a dummy variable D (Note: potential issues with using a dummy variable in the first stage of the regression as discussed below). Thus, the form of the first stage is the reduced form equation:

$$D_{ij} = \gamma_{1ij} * z_{ij} + \omega_{ij}$$

Next, for the second stage of the regression, we will regress the results from the first stage on real revenue in 2016 dollars:

$$R_{ij} = \alpha + \beta_{1ij} * (\gamma_{1ij} * z_{ij}) + \epsilon$$

In practice, we will not be estimating these regressions manually in two stages, but rather using the R-package *AER* which has a built in 2SLS regression function, *ivreg*. In the second stage of the regression, we will include vectors of market and stadium controls as well to ensure that the postseason coefficients are not mistakenly capturing too much of the variation from other effects.

Discontinuity via Data Subsets

One final way we might be able to tease out causality is to look at specific subsets of data for which we might be able to better identify the causal relationship between going to the postseason and increased revenue. The methodology here would be to limit our sample to only those teams within a given range of the average wins it took to get to the postseason for their league (National League or American League) lagged one year. Thus, we will be looking at how the revenues changed for those teams in the current year that were close to making it to the postseason in the previous year. By limiting the data to only these teams, we would be able to isolate the randomness that exists between a team that barely missed the postseason, and a team that went to the postseason in a given year. There are two main downsides to this approach:

1) By limiting the teams to only those teams within a certain range of the average wins to make it to the postseason, we will be significantly lowering the amount of data we will be using, which will no doubt lower the statistical validity of our analysis.

2) We will need to ensure balance on either side of the discontinuity. In other words, we will need to ensure that there are no biases in the samples of teams that were slightly above or slightly below the average number of wins to make it to the postseason. If there are biases in these samples, then we will not be able to trust our final result.

IV. Data Description

Broadly, our data set consists of the 30 MLB teams' yearly records going back to 1993, which comes out to 740 observations. All performance data was gathered from Baseball Reference. Not all teams were part of the data set for the entire time frame of the observations. The Washington Nationals officially began play in 2005 after relocating from Montreal, and the Arizona Diamondbacks joined the league in 1998. We supplemented the seasonal win-loss records and team performance metrics from Baseball Reference with game level attendance data also gathered from Baseball Reference. Additionally, in terms of baseball specific metrics we gathered the Relative Power Index from ESPN, the Strength of Schedule metric from ESPN, the Strength of Schedule metric from Baseball Reference, and we computed our own Opponent Win Percentage from the game level data using the Pythagorean formula of wins. For team specific stadium data (capacity, date built), we used Wikipedia entries for the stadiums. While broadly speaking, Wikipedia would not be a go to source for highly trusted information, based on spot checks, it worked well as a means of gathering team specific stadium data in a centralized location. Finally, for other financial data related to MLB including the Fan Cost Index data, we used a database put together by Rodney Ford of the University of Michigan. The Fan Cost Index, the price of four average tickets, two small beers, four small sodas, four hot dogs, parking for one car, two game programs and two adult-sized caps, is calculated in nominal terms, so we adjusted the index based on 2016 dollars in order to get Real 2016 dollars as the constant value within our data set.

For market level data, we primarily used the Census for population estimates by metropolitan statistical area and data from the St. Louis Fed for per capita income by metropolitan statistical area. For our population data, there were a few issues we would like to point out:

1) The metropolitan statistical areas changed between 1990 and 2017. In order to address this challenge, we used the relevant metropolitan statistical area for the city most associated with the team for a given set of Census tables.

2) The Census only had point in time data for 1990 and 2000, thus we assumed the growth rate over the 10 years was constant. We made the same assumption to estimate populations from 2000 and 2010.

MLB Revenue Sharing

Part of the impetus for exploring attendance revenue versus diversified revenue streams resulted from the difficulty in obtaining dependable figures for each franchise that incorporated adjustments from MLB's revenue sharing program. Baseball initially instituted the revenue sharing concept in 1997 before restructuring and expanding the program in 2002 in an effort to promote on-field parity. Under the 2002 plan, teams pay 31% of net local revenue into a collective fund where the money is then distributed evenly to each team [13]. Teams also receive payouts from Major League Baseball's central fund according to how much revenue they generate. Smaller franchises such as the Tampa Bay Devil Rays and Pittsburgh Pirates benefit from the arrangement and receive additional income, whereas storied franchises such as the Boston Red Sox and New York Yankees are net payers into the system [13]. While the intent is for net receivers to utilize annual cash windfalls to improve the on-field product offered to fans, there are perpetual disagreements about whether sufficient reinvestment occurs. In this year's stagnant free agent market, The Players' Association filed formal grievances against the Oakland Athletics, Miami Marlins, Pittsburgh Pirates, and Tampa Bay Devils Rays [12]. For the purposes of this paper, a thorough forensic accounting analysis of 30 private entities proved unrealistic, as attempts to incorporate revenue sharing figures would have been speculative at best. Exploring revenue derived from regular season attendance and its impact on winning allowed us to maintain the integrity of the data by utilizing figures and indices that can be properly verified.

Explanation of regression variables

Dependent variable

- ❑ Revenue: Is calculated by adjusting the Fan Cost Index to real 2016 dollars, dividing by 4 (to get the average price per person) and multiplying the result by the yearly attendance at the team's stadium.

Independent variables

- ❑ Stadium age (stad_age_1 to stad_age 10): A set of ten binary variables to test effect of stadium age on revenue. stad_age_X equals 1 if the team's stadium is X years old; equals 0 otherwise. We used binary variables for each year to be able to detect in exactly which year this variable becomes insignificant.
- ❑ Wins (LY1_wins to LY5_wins): 5 integer variables included to test effect of number of regular season wins on revenue. LYX_playoffs represents the number of regular season games the team won X years ago.
- ❑ Playoff participation (LY1_playoffs to LY5_playoffs): A set of 5 binary variables included to test effect of playoff participation on revenue. LYX_playoffs equals 1 if the team made it into the playoffs X years ago, equals 0 otherwise. We used binary variables for each different year to be able to detect in exactly which year this variable becomes insignificant.
- ❑ World Series Wins (LY1_WS to LY5_WS): A set of 5 binary variables included to test effect of world series win on revenue. LYX_WS equals 1 if the team won the world series X years ago, equals 0 otherwise. We used binary variables for each different year to be able to detect in exactly which year this variable becomes insignificant.

- ❑ Per Capita Income: Takes on integer values and ranges from $36,603 - $61,850 ($ = USD) and is then consumer price index adjusted to real 2016 dollars. Data acquired from St. Louis Fed
- ❑ Population: Takes on integer values from 1,518,265 to 19,838,284. This data is also acquired from the St. Louis Fed.

Removing Observations

Most of our data is appropriate to include in our study; however, one concern we had was including observations from 1994 and 1995, as these would be significantly impacted by the players' strike. As Figure 4 demonstrates, there was a marked decrease in attendance during these years, mainly due to the truncated season in 1994. In fact, attendance did not return back to its pre-strike levels until the 1998 season.

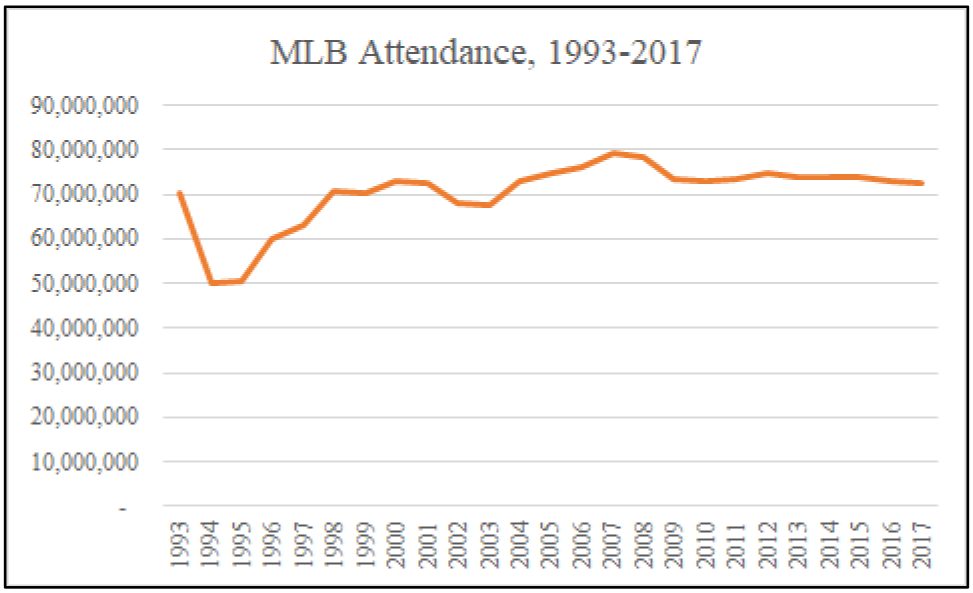


Figure 4: MLB attendance from 1993 - 2017

For this reason, we decided to exclude the seasons from 1993 to 1995. Further, data availability impacted our ability to locate reliable Fan Cost Index numbers for the 2017 season, and so we will be excluding that season as well. This leaves us with 626 observations in our data set out of the original 740. Finally, we removed the two Canadian teams in our data set, the Montreal Expos and the Toronto Blue Jays, as we felt the currency fluctuations and differences in data sources for population and revenue would add more noise rather than signal to our analysis. Thus, we felt it best to remove the respective franchises from our analysis. This leaves us with 596 observations.

These data have a few limitations that were unavoidable. Firstly, the revenue is focused on regular season games and concessions only. This does not allow us to take into account revenues from TV deals, other events at stadiums, and other sponsorship deals which would likely increase the impact of winning in general. Secondly, we rely on the Fan Cost Index to estimate revenue, which may be skewed relative to the actual revenue generation for each team. While this number does vary from team to team, it assumes that for every game, every attendee pays for the basket of goods included in the metric (i.e. hot dog, beer, coke, parking, hat, tickets). Conversely, we believe that a strength of our paper is the transparency of the numbers we have used and the ease with which one could calculate these numbers themselves. Further, focusing on only regular season games creates a cleaner focus on pricing power and increased fan enthusiasm as we ignore the inevitable revenues that would come from 1 - 11 additional home playoff games. Future studies may attempt to update any of these calculations depending on increased data availability, and we certainly feel that a robust effort to incorporate broadcast revenue, MLB's revenue sharing formula, and cost of attendance adjustments by city would offer immense benefits to front offices. The biggest obstacle is obtaining revenues and cost structures from private organizations that are incentivized to hide income in order to avoid paying into the league's revenue sharing program.

V. Data Summary

A brief overview of our data can be found below in Table 2. The table highlights the differences between teams that made the postseason and teams that failed to make the playoffs. At a glance, teams appear to benefit from winning with regards to real revenue generation and yearly attendance. There is also some evidence of pricing power for teams that make the postseason that is reflected in higher Fan Cost Index figures for postseason qualifiers. The table also illustrates the commonly cited 90 win figure required for playoff qualification, as the majority of playoff-bound teams win between 88 and 98 games in the regular season.

Variable	Mean - No Postseason	Std - No Postseason	Mean - Postseason	Std - Postseason
Real Revenue (2016 Dollars)	107,769,153	51,480,031	154,240,239	61,055,537
Yearly Attendance	2,211,581	672,983.9	2,873,258	615,232.2
Real Fan Cost Index (2016)	188.63	44.42	209.27	53.26
Population	5,215,767	4,217,284	6,874,155	5,554,980
Per Capita Income	40,363.43	11,202.8	42,147.17	11,246.51
W	74.94	9.88	93.91	5.62
Stadium Age	25.32	23.11	32.66	30.06

Table 2: Summary Statistics of Playoff and Non-Playoff Participation

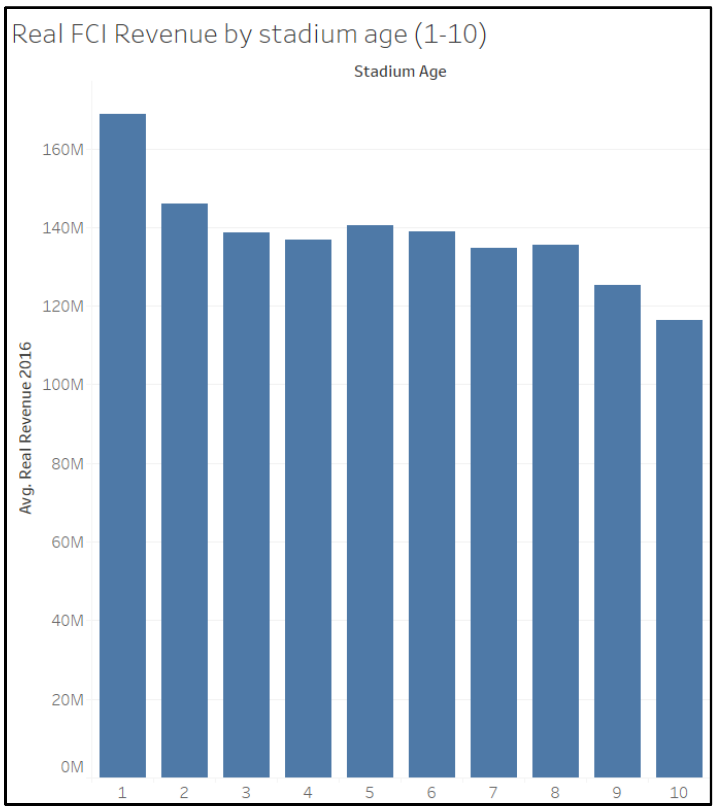


Figure 5: Revenue by stadium age (1 year to 10 years old)

Figure 5 above visually depicts the relationship between revenues and stadium age. Per the base model and previous literature, we expected to see a marked jump in revenue in the year following new stadium construction and to see this effect slowly diminish until the 'honeymoon' period of 10 year ends. We also exclude all stadium renovations from this graph. As an example, The Anaheim Angels' stadium that was built in 1966 underwent significant renovations in 1997 [14]. Given the difficulty in accounting for and differentiating between extensive capital improvements and the various levels of stadium enhancements that always occur from year to year, it is best to exclude everything other than new stadium construction. As expected, for most teams, the revenue is largest in the year following the building of their new stadium. Examples of teams whose revenue show this pattern are the Miami Marlins, Minnesota Twins, Chicago White Sox and San Diego Padres, who moved into Petco Park in 2004 after previously sharing Qualcomm Stadium with the San Diego Chargers of the National Football League (NFL). Our findings confirmed the view that newer stadiums provide a large boost in the first year after construction, and figure 5 shows the same downward glide path posited by other authors.

Figure 6 provides a summary of the number of postseason appearances from 1998-2016. The New York Yankees lead the way with 14 playoff appearances, but the consistency of the St. Louis Cardinals is interesting to note, as it provides visual evidence of "The Cardinal Way" that is often hailed by sports writers and league executives as a model for middle market teams.

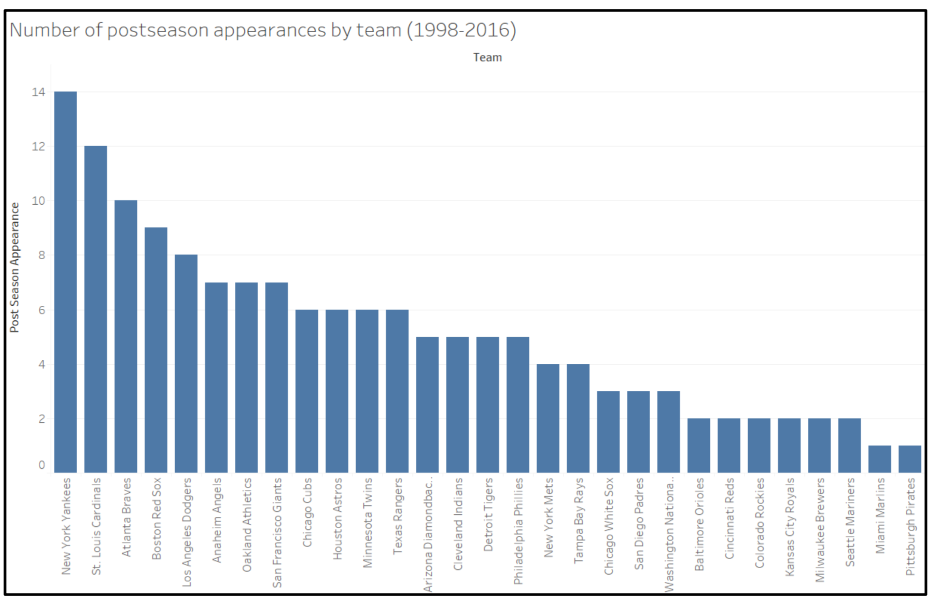


Figure 6: Number of postseason appearances by team

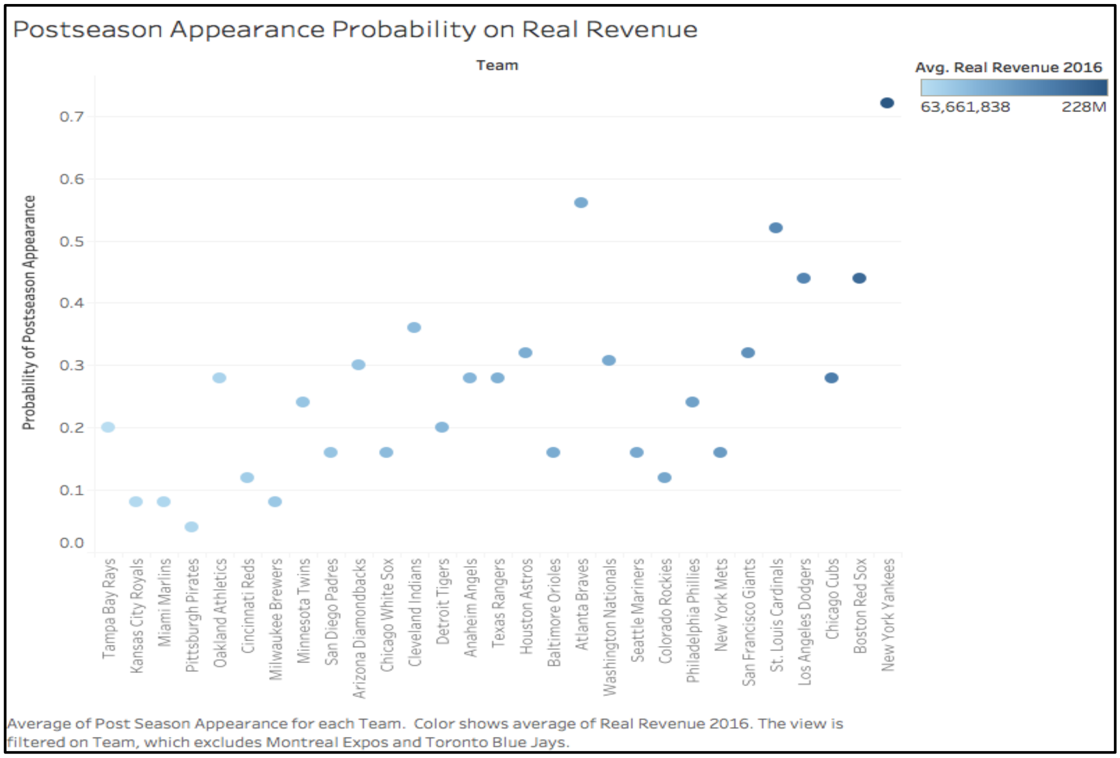


Figure 7: Probability of postseason appearance by team. Darker circles represents higher revenues.

The graph above shows the postseason appearance probability given a team's average revenue. The revenue is derived from the fan cost index, where the fan cost index is divided by four (to calculate the cost to attend a game for one person) and multiplied by the total attendance of the team that season. Both are adjusted based on the consumer price index of 2016. The darker the hue of the circle, the greater the revenue measure is. Unsurprisingly, the New York Yankees' circle is the darkest, and they also have the greatest postseason probability. There is an apparent positive linear relationship between the average real revenue of the team and postseason appearance probability. If the team has a greater revenue, then in a sense, it has the ability to "buy" postseason appearances. As a result, we further analyze the relationship between postseason appearances and revenue in Part VII below, including an exploration of reverse causality with instrumental variables.

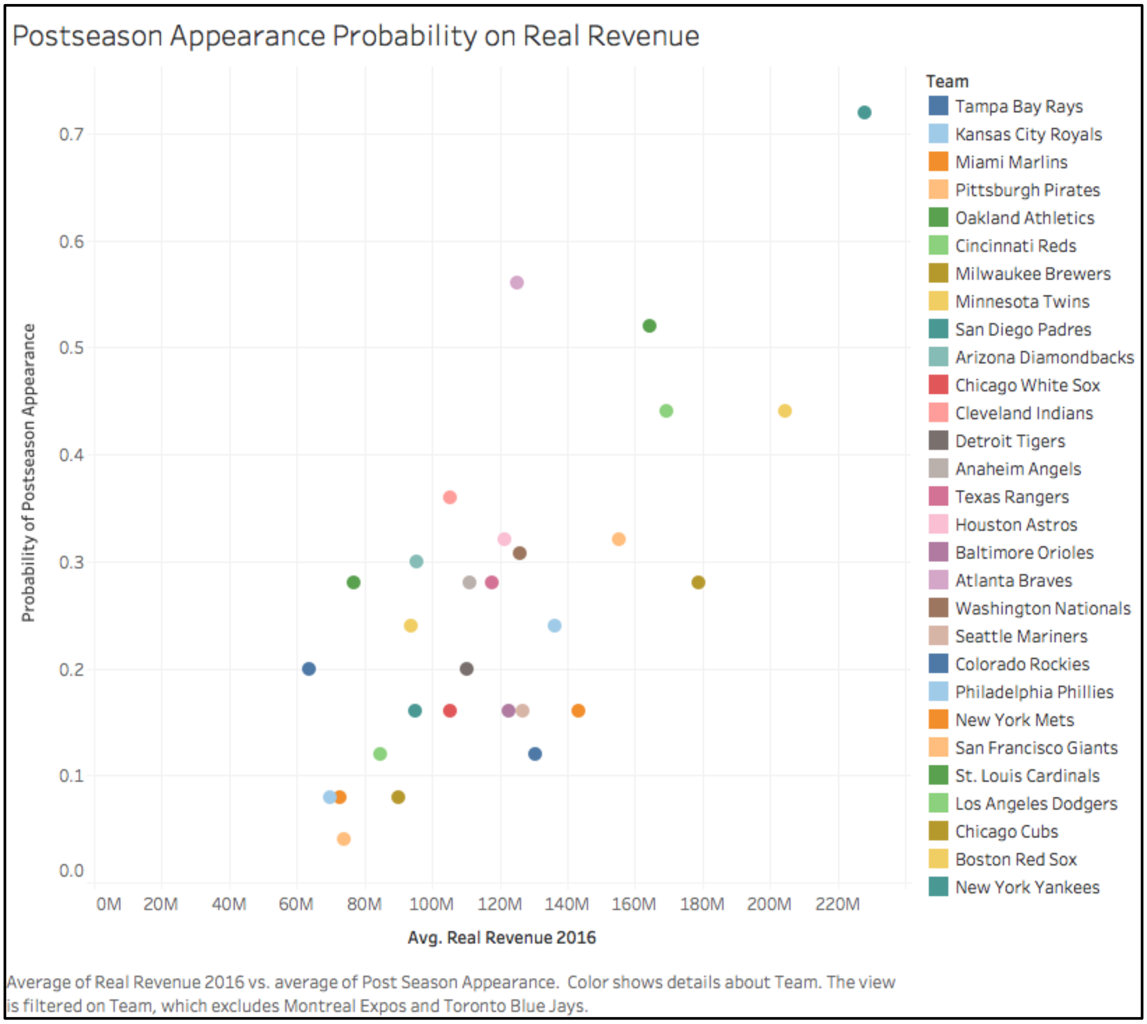


Figure 8: Revenue vs. Probability of Postseason Appearance by team.

Similar to the previous graph, Figure 8 charts the postseason appearance probability given a team's average revenue. However, instead of placing teams on the x axis, real revenue is used. The teams are color coded. It is easier to see the linear relationship between average revenue and probability of postseason appearance in this graph. Two of the more interesting observations in Figure 7 concern the Atlanta Braves and St. Louis Cardinals. Despite revenues that trail major market teams such as the Chicago Cubs, New York Yankees, Boston Red Sox, and Los Angeles Dodgers, the Braves and Cardinals maintained the second and third highest playoff probabilities in our data set, respectively. These are two teams that are consistently praised for their cultures and consistency, and figure 7 provides visual evidence of why the Braves and Cardinals have been held out as franchises to be emulated for teams in smaller markets.

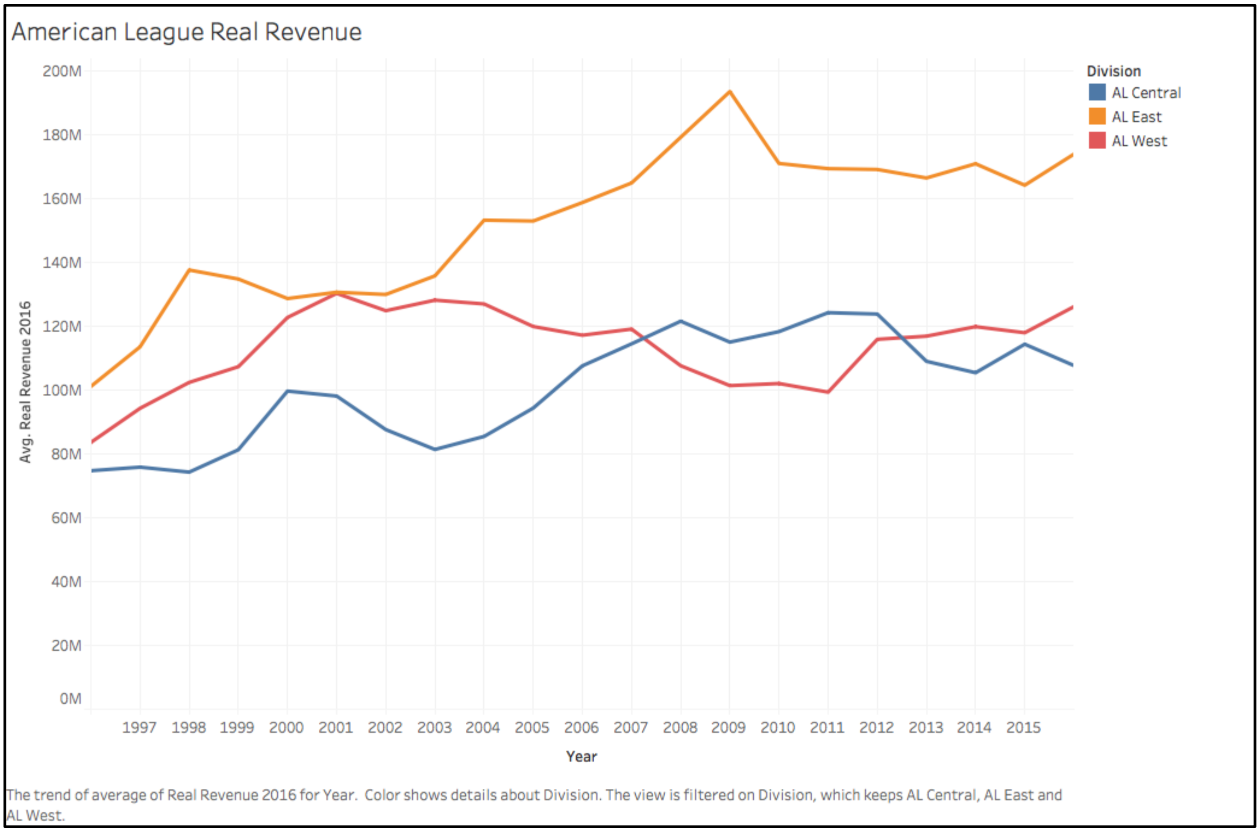


Figure 9: American League Revenue (1996 - 2016)

Another question and concern we had is to see whether divisions in a league have an impact on revenue throughout the years. The graphs above display the trend for the American league. It is apparent that the American League East has consistently generated more attendance revenue than the rest of the league. Not surprisingly, the New York Yankees and Boston Red Sox reside in the American League East. The jump in 2009 is noticeable and coincides with the opening of the new Yankee Stadium in April of the same year. Somewhat surprisingly, the AL West saw a steady decline in revenues from 2001 until 2011. While revenues have gradually recovered and trend upward over the course of our time series, the persistence of stagnant revenues matches the narratives that have continually plagued the division. The Oakland A's and Texas Rangers have both been mentioned in the same breath as relocation on numerous occasions, and their struggles to secure new stadiums are well-documented. When these facts are combined with the limited drawing power of the Angels, it becomes apparent that the division lacks the "anchor tenants", or historically prestigious franchises, that can be found in divisions such as the AL East and NL Central. Similar problems exist in the AL Central, as there is a clear divergence between the AL East and the the other two divisions in the American league.

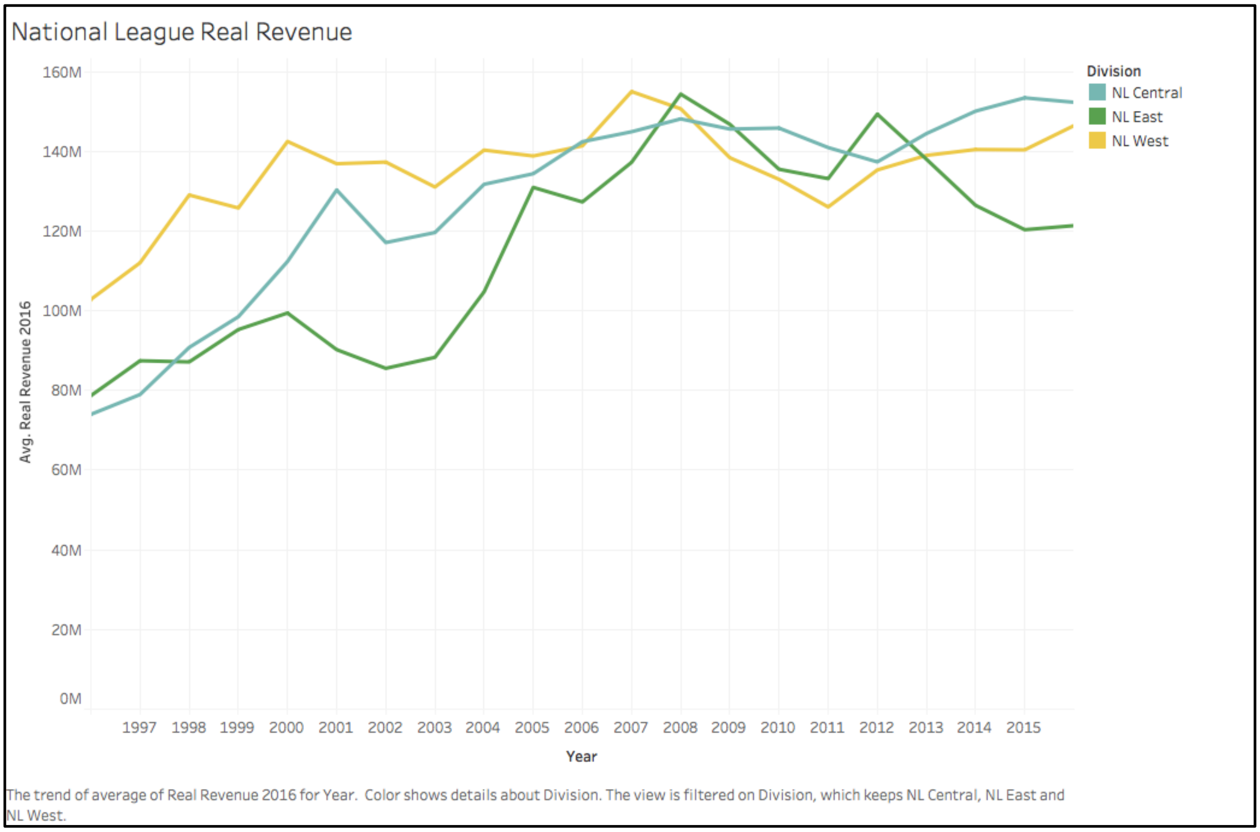


Figure 10: National League Revenue (1996 - 2016)

Figure 10 provides insight into divisional breakdowns in the National League. The sharp divergence in fortunes that is visible in the American League data is absent in the National League. Generally speaking, all National League divisional boats have risen with the tides of time. Part of this can probably be partially attributed to the aforementioned "anchor tenant" effect. Whereas the AL West and AL Central both have franchises located in two-team markets (The Oakland A's and the Chicago White Sox) that have historically trailed their NL counterparts in fan engagement and attendance, The NL West and NL Central are both well supported by the presence of historical heavyweights such as the Los Angeles Dodgers, Chicago Cubs, and San Francisco Giants. Parity in the NL Central is further supported by the historical success of the St. Louis Cardinals. Stadium construction trends in Major League Baseball have also favored the National League Central and West divisions over the American League Central and West. 80% of teams in these two National League divisions have opened new stadiums over the course of our data collection period, and the only older stadiums are Dodger Stadium and

Wrigley Field. Only 50% of American League teams in these two divisions have opened new stadiums in the same time period.

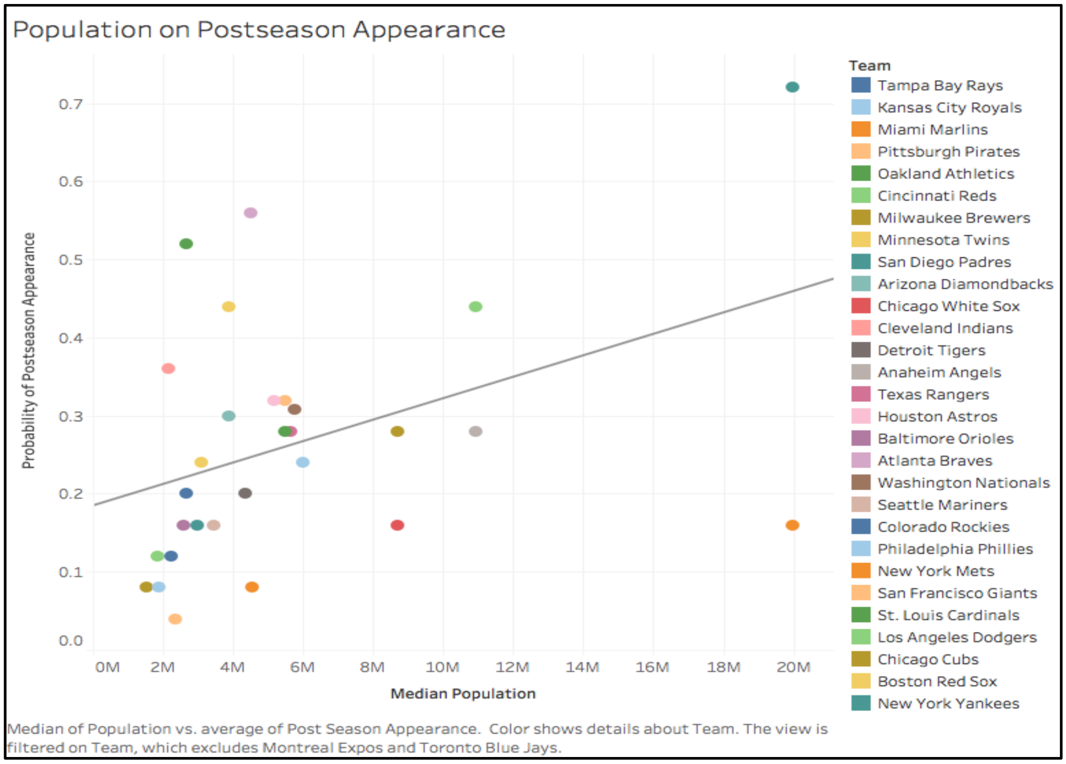


Figure 11: Probability of Postseason Appearance vs. Median Population

There is some evidence of a relationship between population and postseason appearances, but being located in a large metropolitan area does not guarantee success. The New York Mets have been unable to replicate the prolific results of the Yankees, and have not benefited from a large potential fan base in the ways that one might assume. Even with access to larger markets, teams may find it difficult to compete for wallet share in cities where consumers have the option of attending events offered by other professional sports leagues and non-sports related entertainment entities.

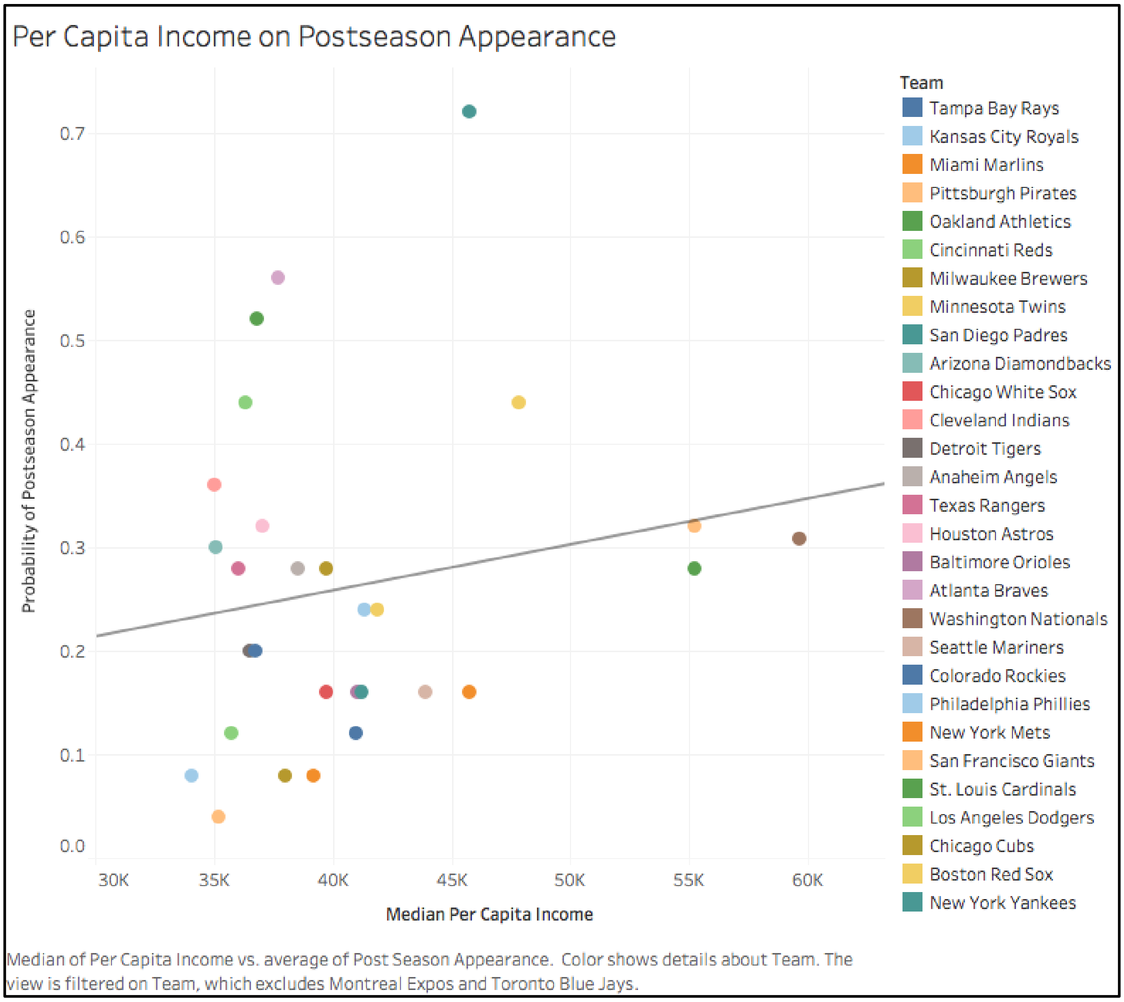


Figure 12:Probability of Postseason Appearance vs. Median Per Capita Income of home city for all teams

Figure 12 shows the relationship between per capita income and postseason appearances. There is a somewhat positive relationship between the two variables, but the predictive power of market size and consumer wealth in determining playoff participation is lower in baseball than one might expect. Part of this may be a result of divisional strength and structure. Teams with similar geographic and economic characteristics are grouped together as a result of Major League Baseball's East/Central/West alignment. With the wild card spots, divisions will typically have at most two teams make the playoffs. The somewhat fortuitous nature of baseball and lack of variance compared to other sports discussed in class may result in larger market teams and teams from wealthier cities missing the playoffs more often than comparable teams in other sports. This has the effect of potentially degrading the effect of per capita income on playoff participation.

VI. Analysis / Hypothesis Testing / Diagnostics

Findings - Fixed Effects

For our fixed effects methodology, we looked at a number of models with different lags to understand the persistence of the impact of postseason appearances. Overall, we looked at 5 different models while steadily increasing the number of lags from one model to the next in order to better understand this relationship. The model that fit the data best including lags going back 5 years for all our winning metrics (wins, postseason appearance, and world series win). Further, we also lagged our stadium age variable; however, in our model that lags 5 years for wins, we went further for stadium age and lagged up to 7 years, given the statistical significance of the additional years of lags for that variable. We also ensured that each of our regressions included only the observations that were relevant for the regression with 5 lags (529 observations).

	1 Lag	2 Lags	3 Lags	4 Lags	5 Lags
Observations	529	529	529	529	529
R2	.799	.825	.840	.847	.859
Adjusted R2	.776	.804	.819	.826	.838
F-Statistic	35.52	38.93	40.12	39.54	39.36

Table 3: Lagged Model Comparison

To determine whether the 5 Lag model was the best of the 5 models, we ran a series of ANOVA tests comparing the 5 Lag model to each of the previous models. In each case, the result was that each of the terms in the 5 Lag model was statistically different from zero, and as a result, the additional lags explained a significant relevant amount of additional information. Within our 5 Lag model, we detected a significant and positive impact for teams that had gone to the postseason in at least 1 of the previous 5 years. We saw a very weak, positive impact of going to the postseason in the 5th year lag with a p-value of .12, but strong positive relationship for the first four years of postseason appearances with p-values all below .05.

	Estimate	Standard Error	T-value	Pr(>|t|)
Postseason 1 Lag	7,418,701	3,533,850	2.099	0.036**
Postseason 2 Lag	12,151,280	3,554,676	3.418	0.001***
Postseason 3 Lag	12,084,840	3,556,481	3.398	0.001***
Postseason 4 Lag	8,124,453	3,616,785	2.246	0.025**
Postseason 5 Lag	5,534,836	3,629,480	1.525	0.128

*p<0.1; **p<0.05; ***p<0.01

Table 4:Lagged Model Estimations

Previous literature states a finding that the positive impact from a postseason appearance degrades over time, 75% in the second year, and 50% in the third, and 30% for the fourth [3]. To determine whether or not such a change in impact occurs, we conducted a Wald Test on the coefficients. The results demonstrated that there is no statistical difference in the impact over the five year span with a Wald Statistic of 3.0 with a p-value of 0.55. Thus, while we can say that there is a positive impact of going to the postseason, we cannot say whether or not that effect deteriorates over time in the way that previous studies have shown.

While our main objective was to look at the impact going to the postseason has on subsequent years' revenues for a team, our fixed effects regression also had interesting findings around stadium age, team fixed effects, and year fixed effects.

We also found a stadium honeymoon effect that Nate Silver and others found; however, we can also say that this effect varies over time, likely due to a decrease in interest in the new stadium as it ages. To prove this we also conducted a Wald Test on the stadium age coefficients to determine dissimilarity, and found a Wald Statistic of 38.8. Further, all of these coefficients were significant with p-values well below .01.

	Estimate	Standard Error	T-value	Pr(>|t|)
Stadium 1 Year Old	63,515,270	7,135,950	8.901	0.000***
Stadium 2 Years Old	36,822,680	6,601,190	5.578	0.000***
Stadium 3 Years Old	28,663,030	6,403,822	4.476	0.000***
Stadium 4 Years Old	25,300,570	6,177,340	4.096	0.000***
Stadium 5 Years Old	20,673,400	5,984,434	3.455	0.001***
Stadium 6 Years Old	17,877,680	5,832,828	3.065	0.002***
Stadium 7 Years Old	15,273,900	5,513,030	2.771	0.006***

*p<0.1; **p<0.05; ***p<0.01

Table 5:Stadium Honeymoon Effect Evaluation

Interestingly, we also found a statistical difference between the coefficients for the team fixed effects with a Wald Statistic of 911.3. The reference team for our fixed effects was the Anaheim Angels, thus all the coefficients in the table below should be thought of as relative to their fixed effects. For ease of showing the data, we are limiting the table to the teams with the largest and smallest coefficients for fixed effects.

	Estimate	Standard Error	T-value	Pr(>|t|)
Chicago Cubs	93,983,620.00	9,884,049.00	9.508615	0.00***
Boston Red Sox	92,506,410.00	16,577,550.00	5.580224	0.00***
St. Louis Cardinals	68,475,230.00	19,926,420.00	3.436404	0.00***
Miami Marlins	(16,846,430.00)	15,796,060.00	-1.0665	0.29
New York Mets	(21,895,530.00)	20,334,470.00	-1.07677	0.28
Oakland Athletics	(70,500,310.00)	16,107,180.00	-4.37695	0.00***

*p<0.1; **p<0.05; ***p<0.01

Table 6:Team summaries relative to Anaheim Angels

While the Marlins and Mets may have relatively large p-values, one should remember that this is only relative to the Angels. Thus, while we cannot say their fixed effects are larger or smaller than the Angels, based on our Wald Test, we can say that there are large differences between the fixed effects for these teams. It should not be all that surprising to see the relative rankings of these teams. The Cubs, Red Sox, and Cardinals are all considered to have incredibly loyal fan bases, despite century long dry spells. One also shouldn't be surprised with the teams that find themselves with the smallest coefficients. The A's and Mets are the "smaller" teams in cities with multiple markets (the Bay Area and New York respectively), and the Marlins have such a non-existent fan base that the Onion has written multiple satire pieces on the team [8].
Finally, in terms of the year fixed effects, we do see a downward trend over time. We tested this trend to see whether or not we can say that there are differences between the various years, and the resulting Wald Statistic (38.9) does lead us to believe there is significance to this trend.

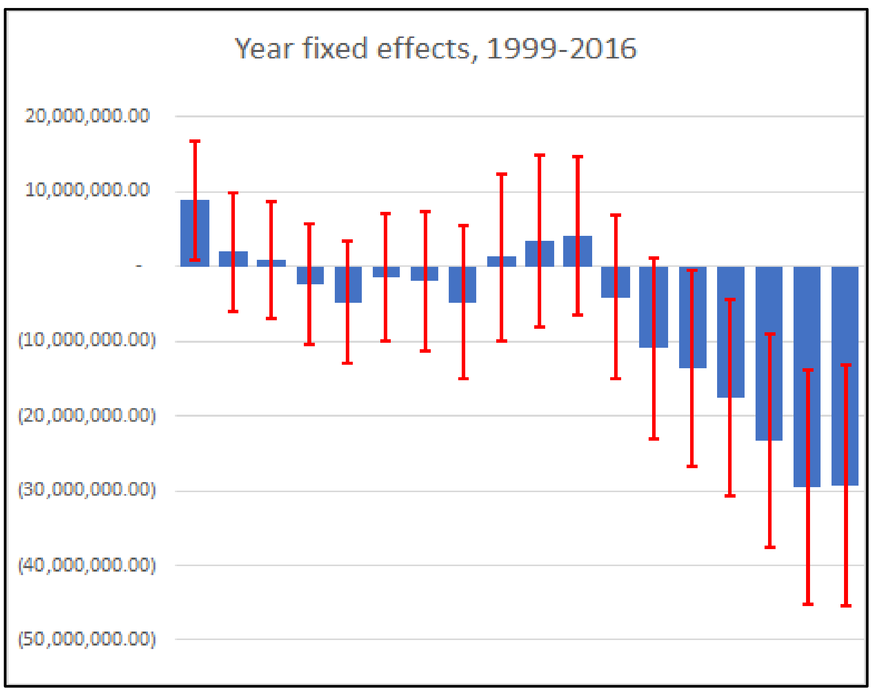


Figure 13: Year fixed effects with standard errors (red)

Explanations for this trend could vary. One study by McKinsey and Co. has found that 20% of Generation X considers themselves "committed" MLB fans relative to only 16% of Millenials [9]. Certainly, the MLB as well as other major sports in the U.S. have been experiencing cyclical declines in viewership as other forms of entertainment vie for the ever shortening attention spans of consumers. While our data do not help us to identify specifically what is causing this decline, it is apparent that there are statistically relevant, negative trends impacting revenue in the MLB.

While the findings from the fixed effects, and stadium honeymoon impact are interesting, the shadow of reverse-causality still looms over our findings for the impact of going to the postseason on subsequent years' revenues.

Forecast Diagnostic - Fixed Effects

One way for us to look at whether including previous performance (playoffs and otherwise) adds information to our model and tells us something important about future revenues is to fit two models on a portion of the data, one with past performance and one without, and forecast these models onto the remaining data. We can then compare these forecasts to understand which of the two models helps to explain the most variation out of sample.

To do this, we split the data into 1998-2010, and 2011-2016. We then fit one model with the past performance (wins, playoffs, and world series wins) and one without past performance. Both models include the year and team fixed effects that we include in our initial fixed effects regression. The base models for each of these forecasts, both explain the data quite well, while the F Statistic for the model with past performance might lead us to believe this explains more of the variation; however, this is not conclusive regarding the superiority of the model with past performance.

Regression	**Adj. R2**	**F Stat**	**P-Value**	**MSE**
Lags (1998-2010)	.841	29.9	2.2e-16	23,269,280
No Lags (1998-2010)	.791	27.7	2.2e-16	26,743,202

We then take these two models, and use them to forecast revenues out of sample. One wrinkle with this forecast is fitting out of sample year fixed effects. To do this, we take the year fixed effects for each model, and fit regress the year on their coefficients to get a baseline for how the coefficients are changing year over year in each model. This then allows us to forecast out the year fixed effects for each forecast. The forecasted fixed effects are in grey below.

Year	No Lags	Lags
2000	(7,282,706)	(6,031,893)
2001	(8,105,070)	(7,413,893)
2002	(12,130,215)	(10,915,357)
2003	(19,680,277)	(16,786,753)
2004	(23,568,697)	(17,937,563)
2005	(28,223,284)	(21,568,525)
2006	(39,806,952)	(30,208,699)
2007	(39,446,235)	(28,256,086)
2008	(39,705,418)	(27,948,833)
2009	(32,087,507)	(22,160,911)
2010	(41,940,672)	(31,505,097)
2011	(45,908,835)	(34,327,974)
2012	(49,876,998)	(37,150,850)
2013	(53,845,160)	(39,973,727)
2014	(57,813,323)	(42,796,604)
2015	(61,781,486)	(45,619,481)
2016	(65,749,649)	(48,442,358)

Table 7: Forecasted fixed effects

Using this method to forecast, we see an MSE of 20,727,491 for the forecast that includes past performance, and an MSE of 22,407,119 for the model that does not include past performance. This lends further evidence to the hypothesis that past performance going back at least 5 years does have a detectable impact on revenue.

Findings - Instrumental Variables

For our instrumental variables regression, as discussed, we used the Baseball Reference Strength of Schedule variable as the instrument, as that was the best predictor of postseason appearance within a given year. For the instrumental variables regression, we used similar controls for market, stadium, team fixed effects, and year fixed effects, except we only use postseason appearance as the winning metric as we could not find enough instruments to include more winning variables.

Unfortunately, we cannot say that appearing in the postseason has a positive impact on subsequent year's revenues. The lowest p-value we get is .12, and none of the playoff coefficients can be said to be statistically different from 0.

	Estimate	Standard Error	T-value	Pr(>|t|)
Postseason 1 Lag	16,708,273	21,560,750	0.7749394	0.44
Postseason 2 Lag	28,354,994	25,631,020	1.10627646	0.27
Postseason 3 Lag	10,990,456	25,789,230	0.42616456	0.67
Postseason 4 Lag	6,519,474	26,335,420	0.2475554	0.80
Postseason 5 Lag	37,601,958	25,029,450	1.50230848	0.13

*p<0.1; **p<0.05; ***p<0.01

Table 8:Instrumental Variable Lags

If the instrument that we used was stronger, we could use this result to, in part, disprove the results of our standard fixed effects regression; however, given the how weak our instrument was, it is hard to say whether the lack of significance in the postseason coefficients is due to the lack of causal relationship between going to the postseason and increased revenues.

Findings - Discontinuity via Data Subsets

The final method we looked at was to subset the data in different ways to take advantage of the potential randomness of teams that were close to going to the postseason but didn't make it, as well as teams going to the postseason in one year but not the next year. One major worry about this method is that there still exists persistent differences between the groups in the subsamples. To ensure this is not the case, we ran a series of tests between the postseason cohort and the non-postseason cohort to ensure balance on factors such as city population and per capita income.

For the data subset with teams +/- 5 games of the average number of wins to make it to the postseason within their league, we ran two sample t tests to understand whether differences existed between the types of teams that made the postseason and the types of teams that didn't. We did this to ensure balance within the dataset.

	T-value	Df	Pr(>|t|)
Per Capita Income	0.71739	58.047	0.476
Population	1.179	75.095	0.242

Based on this, we can at least feel comfortable that the there aren't more large, higher income cities in the postseason cohort relative to the non-postseason cohort, which will help us to feel comfortable about the results that we get from the resulting regression.

Unfortunately, the results of the first subset regression do not show a statistically significant relationship between participating in the postseason and an increase in revenue. The p-value for the postseason coefficient is .48, and thus we cannot reject the possibility that the postseason had no impact on revenue. The adjusted R2 for this regression is much lower than our fixed effects regression at .363 relative to .8376. This is no surprise given we have 159 observations in this subset relative to 596 observations in the overall dataset, so we would expect our regression to not capture as much of the signal.

Model Comparisons

Model	Type	R Square	Adjusted R square	AIC
1	Base model	0.427	0.4126	20126.76
2	Fixed Effects - 1998-2016	0.8595	0.8376	19536.83
3	Fixed Effects Forecast (Lags) - 1998-2010	0.8706	0.8415	13111.9
4	Fixed Effects Forecast (No Lags) - 1998-2010	0.8202	0.7906	13198.66
5	Instrument Variables	0.8376	0.814	N/A
6	Discontinuity Regression	0.4967	03638	6104.632

Table 9: Instrumental Variable Lags

VII. Conclusion

In the end, it is hard for us to say with certainty that the results from our fixed effects regression should be trusted. The results from both the instrumental variables and discontinuity regression show insignificant results for our postseason participation coefficients; however, our instruments are also relatively weak with an F statistic of 7.369 for the first stage. Further, because we focused on only those teams that were +/- 5 games of the playoffs for our discontinuity regression, we severely limited the number of observations which would no doubt lower the statistical significance of any findings. At the very least, when we look at the forecasts of 2011-2016 with and without the performance lags, we can say that past performance does seem to have a significant impact on the variability of year to year revenue for teams, and this past performance trends back at least 5 years for postseason participation.

There are a few ways we would have potentially changed our methodology upon further consideration. Firstly, we think it may have been a better strategy to balance the observations in our instrumental variable. By this we mean including the a 50% split between teams that appeared in the postseason, and teams that did not. As it stands, there are far more teams that did not make it to the postseason in the instrumental variables regression than did. We feel that not balancing this may have biased our data. Additionally, we were quite focused on the postseason appearance, but we may have seen better results if we had focused on wins as opposed to postseason appearances.

Our hope is that this paper can serve as a starting point in identifying and testing stronger instrumental variables, or for more novel approaches to discontinuity regressions that can further help to tease out the causal relationship between winning, specifically postseason participation, and revenue. We suggest future research focus on these areas as a means of helping GMs and teams make optimal tradeoffs with regards to their personnel decisions.

Sources

[1] Bradbury, John Charles. 2016. “Winning and Other Determinants of Revenue in North America's Major Professional Sports Leagues.” Faculty Publications. 4011. http://digitalcommons.kennesaw.edu/facpubs/4011

[2] Angrist, Joshua D. Pischke, Jorn-Steffen. 2008. “Mostly Harmless Econometrics: An Empiricist's Companion.”

[3] Gennaro, Vince. 2013. “Diamond Dollars: The Economics of Winning in Baseball”

[4] Athey, Susan. Imbens, Guido W. 2017. “The State of Applied Econometrics: Causality and Policy Evaluation.” *The Journal of Economic Perspectives*. 31(2): 3-32.

[5] Imbens, Guido W. 2014. "Instrumental Variables: An Econometrician's Perspective." *Statistical Science* 29(3): 323-58

[6] Census, 2018. Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2017 - United States -- Metropolitan and Micropolitan Statistical Area; and for Puerto Rico: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

[7] Census, 2000. Metropolitan Area Population Estimates for July 1, 1999 and Population Change for April 1, 1990 to July 1, 1999 (includes April 1, 1990 Population Estimates Base): https://www2.census.gov/programs-surveys/popest/tables/1990-2000/metro/totals/ma-99-01.txt

[8] The Onion. “Florida Marlins Delay Game Until Their Fan Shows Up”. https://sports.theonion.com/florida-marlins-delay-game-until-their-fan-shows-up-1819571683, The Onion. “ Derek Jeter Denies Tanking Allegations After Marlins Field 4 Players”. https://sports.theonion.com/derek-jeter-denies-tanking-allegations-after-marlins-fi-1825823047

[9] Stinger, Dan. “We are wrong about millennial sports fans”. October 2017, https://www.mckinsey.com/industries/media-and-entertainment/our-insights/we-are-wrong-about-millennial-sports-fans

[10] Baseball Between the Numbers: Why Everything You Know About the Game Is Wrong : Nate Silver, Jonah Keri

[11] Jacobson, David. “MLB’s Revenue-Sharing Formula”. 2008. https://www.cbsnews.com/news/mlbs-revenue-sharing-formula/.

[12] Diamond, Jared. *Wall Street Journal*. “MLB Players Union Accuses Four Teams of Mishandling Revenue-Sharing Funds”. February 27, 2018. https://www.wsj.com/articles/mlb-players-union-accuses-four-teams-of-mishandling-revenue-sharing-funds-1519782386.

[13] Fatsis, Stefan. *Wall Street Journal*. “Playing Hardball”. April 28, 2006. https://www.wsj.com/articles/SB114619267334238458.

[14] Major League Baseball. *Angel Stadium of Anaheim History*. https://www.mlb.com/angels/ballpark/history.

[15] Silver, Adam. “Is Alex Rodriguez Overpaid?” *Baseball Between the Numbers* 5: 174-198.

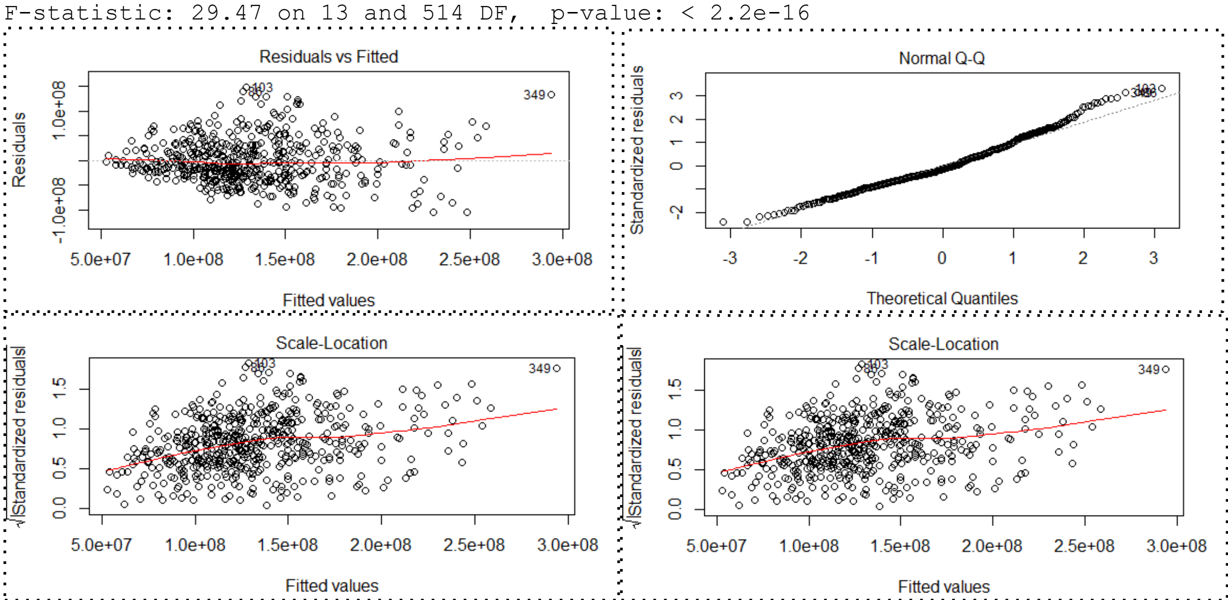
[16] Aschburner, Steve. *Sports Illustrated*. “Inside the GM’s Office: Anatomy of a Deadline Deal.” https://www.si.com/mlb/2008/07/31/aschburner-ryan. July 31, 2008.

[17] Thompson, Wright. *ESPN*. “The Mastermind”. http://www.espn.com/espn/feature/story/_/id/17588399/theo-epstein-mastermind-cubs-season. September 20, 2016.

Appendix

A : Replication of Silver et al Regression Technique

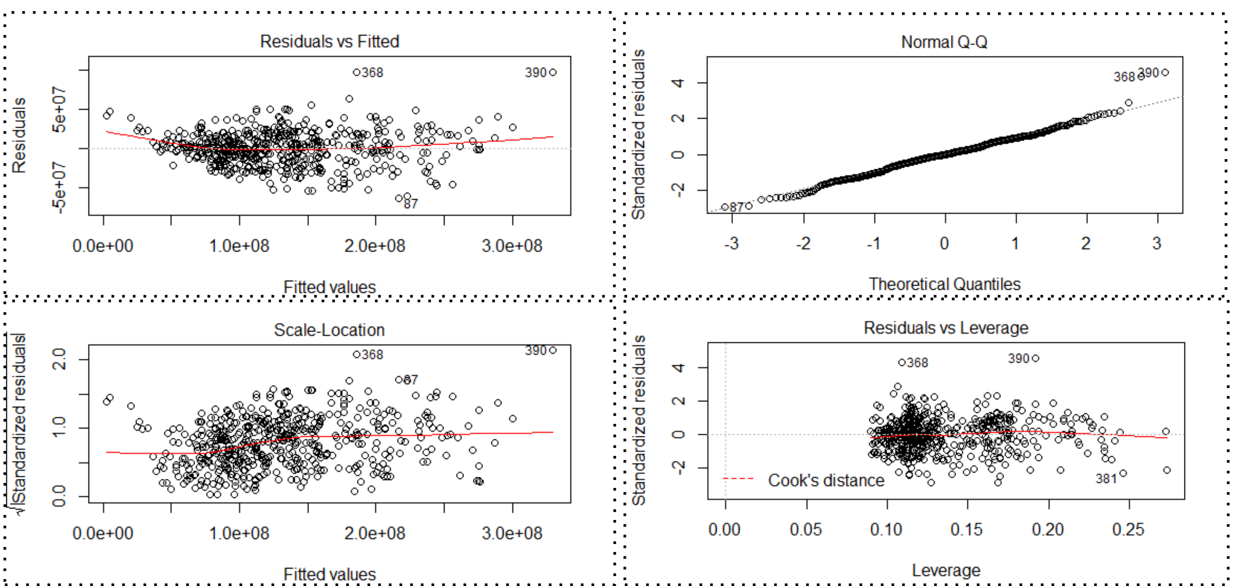
```
Call:
lm(formula = real_revenue_2016 ~ W + +stad_age_1 + stad_age_2 + stad_age_3 +
stad_age_4 + stad_age_5 + stad_age_6 + stad_age_7 + stad_age_8 + stad_age_9 +
stad_age_10 + per_capita_income + population, data = data_right)

Residuals:
      Min         1Q     Median         3Q        Max
-105321535  -29446743   -6711307   26674329  148989200

Coefficients:
                   Estimate Std.     Error      t value     Pr(>|t|)
(Intercept)            -9.397e+07  1.633e+07  -5.755    1.49e-08 ***
W                      1.540e+06  1.768e+05  8.709     < 2e-16 ***
stad_age_1             6.503e+07  1.277e+07  5.092     4.99e-07 ***
stad_age_2             4.058e+07  1.192e+07  3.404     0.000717 ***
stad_age_3             2.472e+07  1.154e+07  2.142     0.032661 *
stad_age_4             2.589e+07  1.121e+07  2.309     0.021318 *
stad_age_5             2.121e+07  1.091e+07  1.944     0.052392 .
stad_age_6             2.118e+07  1.064e+07  1.992     0.046946 *
stad_age_7             1.824e+07  1.015e+07  1.798     0.072782 .
stad_age_8             1.284e+07  1.038e+07  1.237     0.216656
stad_age_9             1.140e+07  1.125e+07  1.013     0.311303
stad_age_10            1.063e+07  1.125e+07  0.945     0.344941
per_capita_income      1.517e+03  2.041e+02  7.434      4.45e-13 ***
population             4.315e+00  4.370e-01    9.872     < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 45150000 on 514 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.427,      Adjusted R-squared:  0.4126
F-statistic: 29.47 on 13 and 514 DF,  p-value: < 2.2e-16
```



B: Ordinary Least Square Regression results

```
Call:
lm(formula = real_revenue_2016 ~ W + LY_wins + LY2_wins + LY3_wins + 
    LY4_wins + LY5_wins + LY_playoffs + LY2_playoffs + LY3_playoffs + 
    LY4_playoffs + LY5_playoffs + LY_WS + LY2_WS + LY3_WS + LY4_WS + 
    LY5_WS + stad_age_1 + stad_age_2 + stad_age_3 + stad_age_4 + 
    stad_age_5 + stad_age_6 + stad_age_7 + per_capita_income + 
    population + Team + Year, data = reg_data)

Residuals:
      Min        1Q    Median        3Q       Max 
-64633185 -13390033    134532  15029392  97717105 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)               -1.794e+08  4.079e+07  -4.399 1.35e-05 ***
W                          5.340e+05  1.140e+05   4.686 3.68e-06 ***
LY_wins                    6.077e+05  1.503e+05   4.043 6.19e-05 ***
LY2_wins                   2.456e+05  1.504e+05   1.633 0.103057    
LY3_wins                   8.332e+04  1.537e+05   0.542 0.588041    
LY4_wins                  -1.878e+04  1.562e+05  -0.120 0.904374    
LY5_wins                   2.669e+05  1.473e+05   1.812 0.070667 .  
LY_playoffs                7.419e+06  3.534e+06   2.099 0.036336 *  
LY2_playoffs               1.215e+07  3.555e+06   3.418 0.000686 ***
LY3_playoffs               1.208e+07  3.556e+06   3.398 0.000738 ***
LY4_playoffs               8.124e+06  3.617e+06   2.246 0.025160 *  
LY5_playoffs               5.535e+06  3.629e+06   1.525 0.127959    
LY_WS                      6.626e+06  6.624e+06   1.000 0.317644    
LY2_WS                     2.514e+06  6.550e+06   0.384 0.701272    
LY3_WS                    -5.050e+06  6.477e+06  -0.780 0.435931    
LY4_WS                    -4.626e+06  6.616e+06  -0.699 0.484797    
LY5_WS                    -4.585e+06  6.722e+06  -0.682 0.495564    
stad_age_1                 6.352e+07  7.136e+06   8.901  < 2e-16 ***
stad_age_2                 3.682e+07  6.601e+06   5.578 4.17e-08 ***
stad_age_3                 2.866e+07  6.404e+06   4.476 9.62e-06 ***
stad_age_4                 2.530e+07  6.177e+06   4.096 4.98e-05 ***
stad_age_5                 2.067e+07  5.984e+06   3.455 0.000603 ***
stad_age_6                 1.788e+07  5.833e+06   3.065 0.002305 ** 
stad_age_7                 1.527e+07  5.513e+06   2.771 0.005825 ** 
per_capita_income          2.582e+03  5.428e+02   4.757 2.64e-06 ***
population                 4.194e+00  2.021e+00   2.076 0.038496 *  
TeamArizona Diamondbacks   1.135e+07  1.931e+07   0.588 0.556909    
TeamAtlanta Braves        -4.103e+06  1.639e+07  -0.250 0.802434    
TeamBaltimore Orioles      3.833e+07  1.952e+07   1.964 0.050193 .  
TeamBoston Red Sox         9.251e+07  1.658e+07   5.580 4.12e-08 ***
TeamChicago Cubs           9.398e+07  9.884e+06   9.509  < 2e-16 ***
TeamChicago White Sox      3.996e+05  9.351e+06   0.043 0.965931    
TeamCincinnati Reds        1.715e+07  2.152e+07   0.797 0.425820    
TeamCleveland Indians      2.304e+07  2.109e+07   1.092 0.275325    
TeamColorado Rockies       4.847e+07  1.996e+07   2.428 0.015573 *  
TeamDetroit Tigers         4.156e+07  1.741e+07   2.387 0.017370 *  
TeamHouston Astros         2.758e+07  1.484e+07   1.859 0.063700 .  
TeamKansas City Royals     2.746e+07  2.172e+07   1.264 0.206790    
TeamLos Angeles Dodgers    5.510e+07  8.139e+06   6.770 3.98e-11 ***
TeamMiami Marlins         -1.685e+07  1.580e+07  -1.066 0.286763    
TeamMilwaukee Brewers      2.778e+07  2.192e+07   1.267 0.205749    
TeamMinnesota Twins        1.184e+06  1.859e+07   0.064 0.949242    
TeamNew York Mets         -2.190e+07  2.033e+07  -1.077 0.282151    
TeamNew York Yankees       2.097e+07  2.095e+07   1.001 0.317268    
TeamOakland Athletics     -7.050e+07  1.611e+07  -4.377 1.49e-05 ***
TeamPhiladelphia Phillies  3.130e+07  1.347e+07   2.323 0.020598 *  
TeamPittsburgh Pirates     1.775e+07  2.086e+07   0.851 0.395364    
TeamSan Diego Padres       1.573e+07  1.898e+07   0.829 0.407730    
TeamSan Francisco Giants   1.208e+07  1.577e+07   0.766 0.443804    
```

```
TeamSeattle Mariners         3.133e+07  1.763e+07   1.777 0.076303 .
TeamSt. Louis Cardinals      6.848e+07  1.993e+07   3.436 0.000643 ***
TeamTampa Bay Rays          -5.752e+06  2.142e+07  -0.269 0.788380
TeamTexas Rangers            2.344e+07  1.401e+07   1.672 0.095138 .
TeamWashington Nationals    -1.614e+07  1.662e+07  -0.971 0.331993
Year1999                     8.831e+06  7.920e+06   1.115 0.265439
Year2000                     1.946e+06  7.950e+06   0.245 0.806717
Year2001                     8.984e+05  7.882e+06   0.114 0.909300
Year2002                    -2.415e+06  8.001e+06  -0.302 0.762876
Year2003                    -4.809e+06  8.093e+06  -0.594 0.552656
Year2004                    -1.448e+06  8.623e+06  -0.168 0.866713
Year2005                    -2.021e+06  9.247e+06  -0.219 0.827114
Year2006                    -4.827e+06  1.028e+07  -0.470 0.638871
Year2007                     1.220e+06  1.108e+07   0.110 0.912372
Year2008                     3.427e+06  1.158e+07   0.296 0.767377
Year2009                     4.015e+06  1.055e+07   0.380 0.703796
Year2010                    -4.165e+06  1.092e+07  -0.381 0.703024
Year2011                    -1.094e+07  1.208e+07  -0.905 0.365750
Year2012                    -1.367e+07  1.309e+07  -1.044 0.297079
Year2013                    -1.756e+07  1.320e+07  -1.331 0.183983
Year2014                    -2.331e+07  1.435e+07  -1.624 0.104993
Year2015                    -2.956e+07  1.559e+07  -1.896 0.058586 .
Year2016                    -2.924e+07  1.616e+07  -1.809 0.071069 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23720000 on 457 degrees of freedom
  (67 observations deleted due to missingness)
Multiple R-squared:  0.8595,	Adjusted R-squared:  0.8376
F-statistic: 39.36 on 71 and 457 DF,  p-value: < 2.2e-16
```



C: Instrument Variable Regression results

Error in UseMethod("logLik") : no applicable method for 'logLik' applied to an object of class "ivreg"

```
Residuals:
      Min        1Q    Median        3Q       Max
-75773995 -14183526    361559  15080041  93366717

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)               -2.269e+07  8.439e+07  -0.269 0.788151
W                          8.280e+05  7.085e+05   1.169 0.243219
LY_playoffs                1.671e+07  2.156e+07   0.775 0.438841
LY2_playoffs               2.835e+07  2.563e+07   1.106 0.269284
LY3_playoffs               1.099e+07  2.579e+07   0.426 0.670221
LY4_playoffs               6.519e+06  2.634e+07   0.248 0.804608
LY5_playoffs               3.760e+07  2.503e+07   1.502 0.133821
stad_age_1                 7.654e+07  9.823e+06   7.793 5.93e-14 ***
stad_age_2                 3.736e+07  9.161e+06   4.078 5.49e-05 ***
stad_age_3                 2.495e+07  7.464e+06   3.343 0.000908 ***
stad_age_4                 1.840e+07  7.623e+06   2.414 0.016250 *
stad_age_5                 1.593e+07  7.078e+06   2.250 0.025013 *
stad_age_6                 1.352e+07  7.202e+06   1.877 0.061281 .
stad_age_7                 1.048e+07  7.036e+06   1.489 0.137166
per_capita_income          1.338e+03  8.865e+02   1.509 0.132132
Year2002                  -1.779e+06  7.437e+06  -0.239 0.811081
Year2003                  -1.600e+06  7.269e+06  -0.220 0.825936
Year2004                   3.446e+06  7.717e+06   0.447 0.655439
Year2005                   7.431e+06  8.241e+06   0.902 0.367752
Year2006                   7.686e+06  9.715e+06   0.791 0.429321
Year2007                   1.652e+07  1.090e+07   1.516 0.130232
Year2008                   2.048e+07  1.139e+07   1.798 0.072938 .
Year2009                   1.731e+07  1.024e+07   1.690 0.091805 .
Year2010                   1.215e+07  1.021e+07   1.189 0.235039
Year2011                   7.603e+06  1.229e+07   0.619 0.536563
Year2012                   8.319e+06  1.374e+07   0.606 0.545083
Year2013                   5.192e+06  1.383e+07   0.375 0.707630
Year2014                   3.220e+06  1.550e+07   0.208 0.835516
Year2015                  -6.787e+05  1.762e+07  -0.039 0.969291
Year2016                   9.983e+05  1.856e+07   0.054 0.957126
TeamArizona Diamondbacks  -4.594e+07  1.677e+07  -2.740 0.006417 **
TeamAtlanta Braves        -6.077e+07  1.676e+07  -3.627 0.000325 ***
TeamBaltimore Orioles     -1.469e+07  1.281e+07  -1.147 0.252122
TeamBoston Red Sox         7.461e+07  1.818e+07   4.105 4.93e-05 ***
TeamChicago Cubs           8.039e+07  1.101e+07   7.303 1.57e-12 ***
TeamChicago White Sox     -1.065e+07  1.035e+07  -1.028 0.304502
TeamCincinnati Reds       -3.029e+07  1.379e+07  -2.197 0.028608 *
TeamCleveland Indians     -3.980e+07  1.181e+07  -3.370 0.000826 ***
TeamColorado Rockies      -1.173e+06  1.285e+07  -0.091 0.927359
TeamDetroit Tigers        -1.618e+06  1.455e+07  -0.111 0.911473
TeamHouston Astros        -8.387e+06  1.270e+07  -0.660 0.509448
TeamKansas City Royals    -2.512e+07  1.780e+07  -1.411 0.158953
TeamLos Angeles Dodgers    4.616e+07  1.192e+07   3.871 0.000127 ***
TeamMiami Marlins         -5.195e+07  1.161e+07  -4.476 9.99e-06 ***
TeamMilwaukee Brewers     -2.067e+07  1.247e+07  -1.658 0.098132 .
TeamMinnesota Twins       -4.662e+07  1.314e+07  -3.549 0.000434 ***
TeamNew York Mets          1.679e+07  1.210e+07   1.388 0.165993
TeamNew York Yankees       4.422e+07  2.911e+07   1.519 0.129471
TeamOakland Athletics     -8.934e+07  2.109e+07  -4.236 2.83e-05 ***
TeamPhiladelphia Phillies  5.677e+06  1.172e+07   0.484 0.628454
TeamPittsburgh Pirates    -3.193e+07  1.532e+07  -2.085 0.037746 *
TeamSan Diego Padres      -3.093e+07  1.156e+07  -2.675 0.007784 **
TeamSan Francisco Giants   8.022e+06  2.112e+07   0.380 0.704204
TeamSeattle Mariners      -2.711e+06  1.095e+07  -0.248 0.804645
TeamSt. Louis Cardinals    9.945e+05  1.782e+07   0.056 0.955510
TeamTampa Bay Rays        -6.693e+07  1.536e+07  -4.358 1.68e-05 ***
TeamTexas Rangers         -1.453e+07  1.063e+07  -1.366 0.172711
```

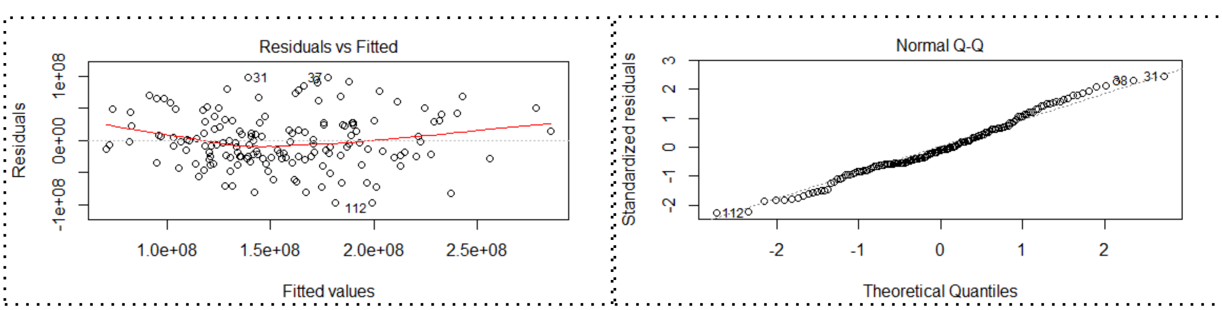
```
TeamWashington Nationals  -3.044e+07  1.616e+07  -1.883 0.060427 .

Diagnostic tests:
                                df1 df2 statistic  p-value
Weak instruments (W)              6 393     3.990 0.000686 ***
Weak instruments (LY_playoffs)    6 393     2.624 0.016591 *
Weak instruments (LY2_playoffs)   6 393     2.012 0.063072 .
Weak instruments (LY3_playoffs)   6 393     2.008 0.063627 .
Weak instruments (LY4_playoffs)   6 393     1.847 0.088819 .
Weak instruments (LY5_playoffs)   6 393     1.601 0.145439
Wu-Hausman                        6 387     0.545 0.773911
Sargan                            0  NA        NA       NA

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25850000 on 393 degrees of freedom
Multiple R-Squared: 0.8376,      Adjusted R-squared: 0.814
Wald test: 35.07 on 57 and 393 DF,  p-value: < 2.2e-16
```

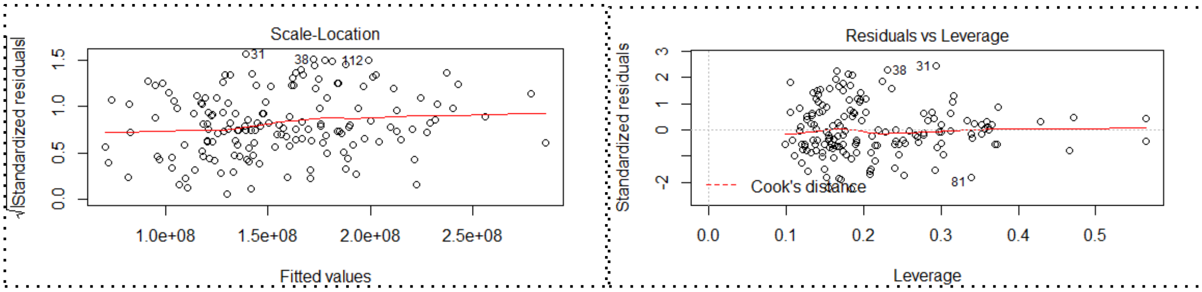
Plot has an error: Error in xy.coords(x, y, xlabel, ylabel, log) : 'x' is a list, but does not have components 'x' and 'y'
4.
stop("'x' is a list, but does not have components 'x' and 'y'")
3.
xy.coords(x, y, xlabel, ylabel, log)
2.
plot.default(iv.reg.real.rev1)
1.
plot(iv.reg.real.rev1)

D: Discontinuity Regression results

```
Call:
lm(formula = real_revenue_2016 ~ W + LY_wins + LY_playoffs +
    LY_WS + stad_age_1 + stad_age_2 + stad_age_3 + stad_age_4 +
    stad_age_5 + stad_age_6 + stad_age_7 + per_capita_income +
    population + Year, data = reg_data_random)
```

```
Residuals:
      Min        1Q    Median        3Q       Max 
-97527300 -25684667  -4298590  27287885  98438455 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -1.767e+07  1.042e+08  -0.170   0.8656    
W                  1.083e+06  4.420e+05   2.450   0.0157 *  
LY_wins           -8.481e+05  1.244e+06  -0.682   0.4968    
LY_playoffs        7.359e+06  1.047e+07   0.703   0.4834    
LY_WS              1.633e+07  1.405e+07   1.162   0.2473    
stad_age_1         4.879e+07  2.996e+07   1.628   0.1059    
stad_age_2         4.841e+07  3.688e+07   1.313   0.1917    
stad_age_3         4.291e+07  2.179e+07   1.969   0.0511 .  
stad_age_4         3.317e+07  2.054e+07   1.615   0.1087    
stad_age_5         2.955e+07  2.130e+07   1.387   0.1678    
stad_age_6         3.380e+07  1.871e+07   1.806   0.0733 .  
stad_age_7         2.587e+07  2.602e+07   0.994   0.3222    
per_capita_income  2.340e+03  5.739e+02   4.078 8.04e-05 ***
population         3.807e+00  7.800e-01   4.881 3.15e-06 ***
Year1997           1.352e+07  3.496e+07   0.387   0.6995    
Year1998          -1.541e+07  3.589e+07  -0.429   0.6684    
Year1999           2.671e+07  4.278e+07   0.624   0.5335    
Year2000           3.142e+07  3.753e+07   0.837   0.4041    
Year2001           4.751e+06  3.488e+07   0.136   0.8919    
Year2002          -4.983e+05  3.663e+07  -0.014   0.9892    
Year2003           3.861e+06  3.971e+07   0.097   0.9227    
Year2004           7.981e+06  4.121e+07   0.194   0.8467    
Year2005           8.036e+06  3.647e+07   0.220   0.8259    
Year2006           2.416e+07  3.653e+07   0.661   0.5096    
Year2007           1.421e+07  3.693e+07   0.385   0.7010    
Year2008           3.377e+07  3.470e+07   0.973   0.3322    
Year2009           2.746e+07  3.653e+07   0.752   0.4537    
Year2010           4.393e+07  3.776e+07   1.163   0.2469    
Year2011           1.441e+07  3.569e+07   0.404   0.6872    
Year2012           2.781e+07  3.755e+07   0.741   0.4603    
Year2013          -1.438e+07  3.667e+07  -0.392   0.6956    
Year2014           2.615e+06  3.643e+07   0.072   0.9429    
Year2015          -2.427e+06  3.728e+07  -0.065   0.9482    
Year2016          -1.556e+06  3.716e+07  -0.042   0.9667    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47590000 on 125 degrees of freedom
Multiple R-squared:  0.4967,	Adjusted R-squared:  0.3638 
F-statistic: 3.738 on 33 and 125 DF,  p-value: 5.196e-08
```

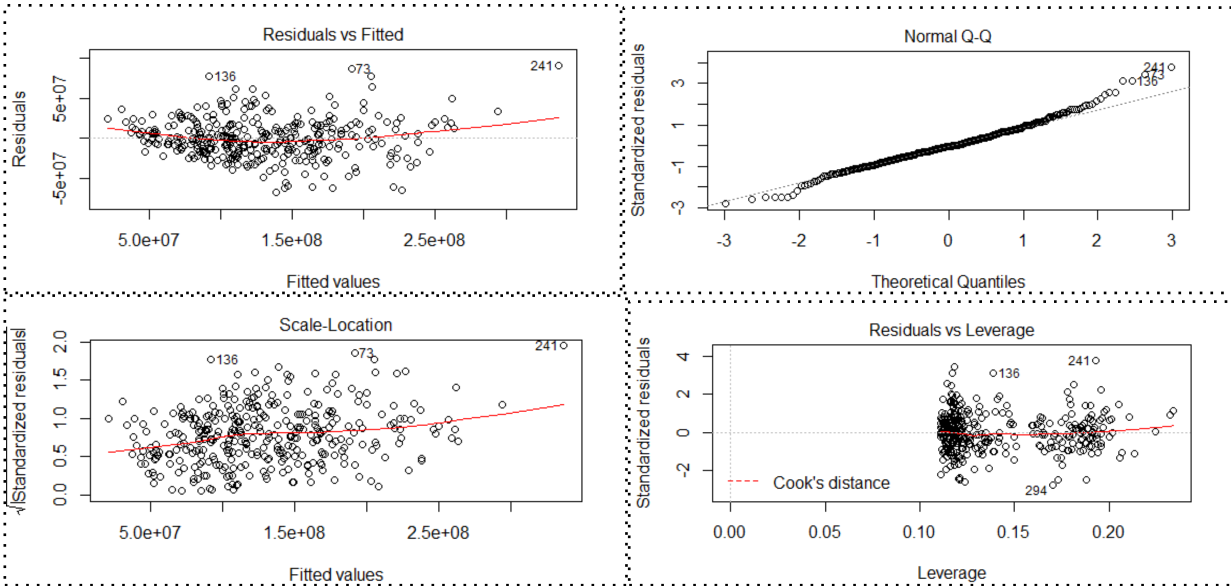




E: OLS Forecast without lags

```
Call:
lm(formula = real_revenue_2016 ~ W + stad_age_1 + stad_age_2 +
    stad_age_3 + stad_age_4 + stad_age_5 + stad_age_6 + stad_age_7 +
    per_capita_income + population + Team + Year, data = data_right_1)

Residuals:
      Min        1Q    Median        3Q       Max
-68182087 -15486682  -1073045  14048947  90984428

Coefficients:
```

```
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)              -2.124e+08  5.140e+07  -4.133 4.64e-05 ***
W                         1.022e+06  1.565e+05   6.532 2.72e-10 ***
stad_age_1                6.565e+07  8.410e+06   7.806 9.63e-14 ***
stad_age_2                3.543e+07  8.101e+06   4.374 1.68e-05 ***
stad_age_3                2.964e+07  8.816e+06   3.362 0.000872 ***
stad_age_4                3.693e+07  8.406e+06   4.393 1.55e-05 ***
stad_age_5                3.410e+07  8.045e+06   4.238 2.99e-05 ***
stad_age_6                2.822e+07  7.750e+06   3.641 0.000319 ***
stad_age_7                2.354e+07  7.234e+06   3.255 0.001264 ** 
per_capita_income         5.488e+03  1.007e+03   5.449 1.05e-07 ***
population                5.066e+00  3.235e+00   1.566 0.118353    
TeamArizona Diamondbacks  4.523e+07  2.741e+07   1.650 0.099965 .  
TeamAtlanta Braves        3.413e+07  2.317e+07   1.473 0.141787    
TeamBaltimore Orioles     5.096e+07  2.779e+07   1.834 0.067685 .  
TeamBoston Red Sox        7.755e+07  2.524e+07   3.073 0.002314 ** 
TeamChicago Cubs          8.056e+07  1.247e+07   6.458 4.19e-10 ***
TeamChicago White Sox     1.403e+06  1.249e+07   0.112 0.910637    
TeamCincinnati Reds       2.031e+07  3.105e+07   0.654 0.513597    
TeamCleveland Indians     6.522e+07  3.015e+07   2.163 0.031304 *  
TeamColorado Rockies      3.784e+07  2.886e+07   1.311 0.190753    
TeamDetroit Tigers        3.208e+07  2.329e+07   1.377 0.169486    
TeamHouston Astros        4.319e+07  2.104e+07   2.052 0.040990 *  
TeamKansas City Royals    2.300e+07  3.090e+07   0.744 0.457246    
TeamLos Angeles Dodgers   7.581e+07  1.066e+07   7.113 8.15e-12 ***
TeamMiami Marlins        -2.294e+07  2.263e+07  -1.014 0.311555    
TeamMilwaukee Brewers     1.451e+07  3.133e+07   0.463 0.643515    
TeamMinnesota Twins      -7.741e+06  2.651e+07  -0.292 0.770503    
TeamNew York Mets        -3.315e+07  3.409e+07  -0.973 0.331553    
TeamNew York Yankees      2.367e+07  3.394e+07   0.697 0.486095    
TeamOakland Athletics    -9.621e+07  2.318e+07  -4.151 4.31e-05 ***
TeamPhiladelphia Phillies 1.288e+07  1.827e+07   0.705 0.481414    
TeamPittsburgh Pirates    1.316e+07  2.941e+07   0.448 0.654782    
TeamSan Diego Padres      2.050e+07  2.695e+07   0.760 0.447597    
TeamSan Francisco Giants -3.104e+07  2.300e+07  -1.350 0.178122    
TeamSeattle Mariners      3.509e+07  2.521e+07   1.392 0.164904    
TeamSt. Louis Cardinals   9.273e+07  2.839e+07   3.266 0.001215 ** 
TeamTampa Bay Rays        2.061e+07  3.053e+07   0.675 0.500246    
TeamTexas Rangers         3.027e+07  1.992e+07   1.519 0.129776    
TeamWashington Nationals -8.531e+07  3.588e+07  -2.378 0.018037 *  
Year1999                  2.499e+06  7.599e+06   0.329 0.742518    
Year2000                 -7.283e+06  8.517e+06  -0.855 0.393192    
Year2001                 -8.105e+06  8.926e+06  -0.908 0.364584    
Year2002                 -1.213e+07  9.105e+06  -1.332 0.183795    
Year2003                 -1.968e+07  9.599e+06  -2.050 0.041194 *  
Year2004                 -2.357e+07  1.094e+07  -2.154 0.032006 *  
Year2005                 -2.822e+07  1.250e+07  -2.258 0.024646 *  
Year2006                 -3.981e+07  1.500e+07  -2.654 0.008372 ** 
Year2007                 -3.945e+07  1.669e+07  -2.363 0.018766 *  
Year2008                 -3.971e+07  1.765e+07  -2.249 0.025205 *  
Year2009                 -3.209e+07  1.547e+07  -2.074 0.038941 *  
Year2010                 -4.194e+07  1.630e+07  -2.573 0.010554 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26740000 on 304 degrees of freedom
Multiple R-squared:  0.8202,     Adjusted R-squared:  0.7906 
F-statistic: 27.73 on 50 and 304 DF,  p-value: < 2.2e-16
```

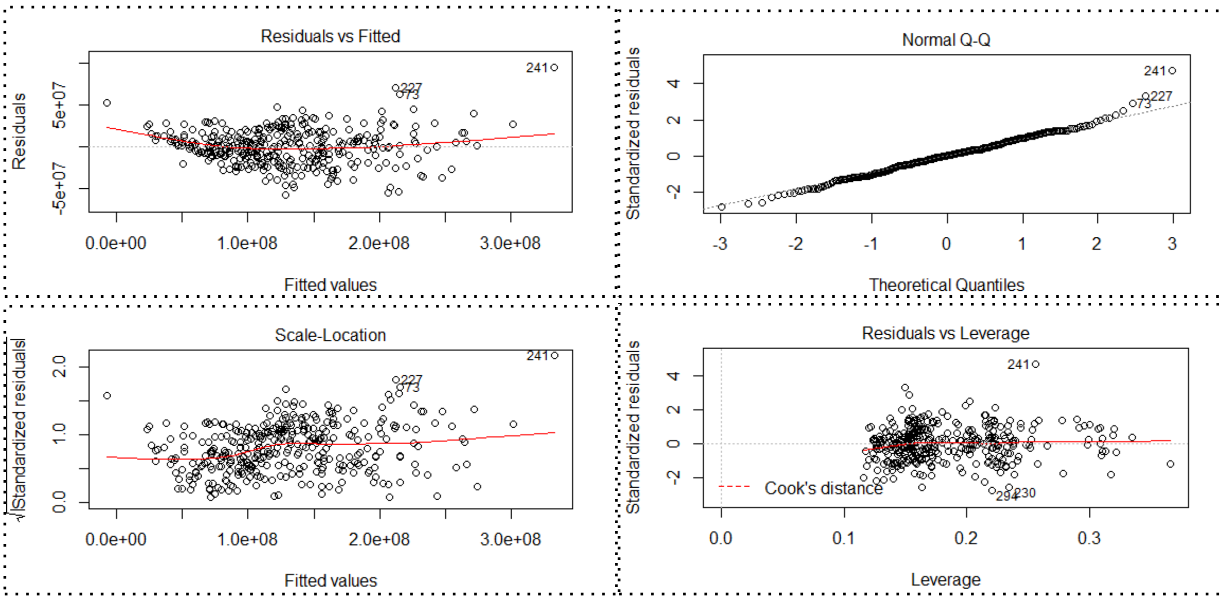


F: OLS Forecast with lags

```
Call:
lm(formula = real_revenue_2016 ~ W + LY_wins + LY2_wins + LY3_wins +
    LY4_wins + LY5_wins + LY_playoffs + LY2_playoffs + LY3_playoffs +
    LY4_playoffs + LY5_playoffs + LY_WS + LY2_WS + LY3_WS + LY4_WS +
    LY5_WS + stad_age_1 + stad_age_2 + stad_age_3 + stad_age_4 +
    stad_age_5 + stad_age_6 + stad_age_7 + per_capita_income +
    population + Team + Year, data = data_right_1)

Residuals:
      Min        1Q    Median        3Q       Max
-57875885 -12632507    327282  13230595  94676629

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)               -2.452e+08  5.250e+07  -4.669 4.63e-06 ***
```

```
W                         7.862e+05  1.428e+05   5.507 8.06e-08 ***
LY_wins                   7.218e+05  1.808e+05   3.992 8.33e-05 ***
LY2_wins                  2.390e+05  1.820e+05   1.314 0.190019
LY3_wins                  9.568e+04  1.918e+05   0.499 0.618303
LY4_wins                 -3.882e+04  1.931e+05  -0.201 0.840786
LY5_wins                  1.372e+05  1.818e+05   0.755 0.450914
LY_playoffs               3.606e+06  4.317e+06   0.835 0.404205
LY2_playoffs              1.291e+07  4.346e+06   2.970 0.003223 **
LY3_playoffs              1.227e+07  4.409e+06   2.783 0.005742 **
LY4_playoffs              6.017e+06  4.507e+06   1.335 0.182932
LY5_playoffs              7.501e+06  4.507e+06   1.664 0.097122 .
LY_WS                     9.981e+06  8.201e+06   1.217 0.224555
LY2_WS                   -3.667e+06  7.834e+06  -0.468 0.640053
LY3_WS                   -1.308e+07  7.783e+06  -1.680 0.093942 .
LY4_WS                   -8.102e+06  8.084e+06  -1.002 0.317056
LY5_WS                   -8.819e+06  8.518e+06  -1.035 0.301365
stad_age_1                6.287e+07  7.424e+06   8.468 1.28e-15 ***
stad_age_2                3.345e+07  7.215e+06   4.636 5.39e-06 ***
stad_age_3                2.767e+07  7.769e+06   3.561 0.000431 ***
stad_age_4                2.949e+07  7.424e+06   3.973 8.99e-05 ***
stad_age_5                2.650e+07  7.105e+06   3.730 0.000231 ***
stad_age_6                1.770e+07  6.921e+06   2.557 0.011059 *
stad_age_7                1.617e+07  6.453e+06   2.506 0.012769 *
per_capita_income         4.629e+03  9.287e+02   4.985 1.07e-06 ***
population                2.834e+00  2.884e+00   0.983 0.326597
TeamArizona Diamondbacks  2.475e+07  2.478e+07   0.999 0.318671
TeamAtlanta Braves       -1.298e+07  2.129e+07  -0.610 0.542613
TeamBaltimore Orioles     4.402e+07  2.456e+07   1.792 0.074106 .
TeamBoston Red Sox        5.538e+07  2.233e+07   2.480 0.013716 *
TeamChicago Cubs          7.892e+07  1.134e+07   6.958 2.31e-11 ***
TeamChicago White Sox    -1.567e+06  1.095e+07  -0.143 0.886271
TeamCincinnati Reds       1.226e+07  2.742e+07   0.447 0.655237
TeamCleveland Indians     2.926e+07  2.692e+07   1.087 0.278032
TeamColorado Rockies      3.283e+07  2.545e+07   1.290 0.198198
TeamDetroit Tigers        3.536e+07  2.088e+07   1.694 0.091361 .
TeamHouston Astros        1.855e+07  1.876e+07   0.989 0.323600
TeamKansas City Royals    2.245e+07  2.751e+07   0.816 0.415233
TeamLos Angeles Dodgers   6.724e+07  9.584e+06   7.015 1.63e-11 ***
TeamMiami Marlins        -2.797e+07  2.012e+07  -1.390 0.165615
TeamMilwaukee Brewers     1.187e+07  2.772e+07   0.428 0.668726
TeamMinnesota Twins      -2.464e+07  2.352e+07  -1.048 0.295688
TeamNew York Mets        -6.983e+06  3.088e+07  -0.226 0.821233
TeamNew York Yankees      1.384e+07  3.024e+07   0.458 0.647472
TeamOakland Athletics    -1.038e+08  2.053e+07  -5.057 7.57e-07 ***
TeamPhiladelphia Phillies 1.292e+07  1.607e+07   0.804 0.421781
TeamPittsburgh Pirates    1.427e+07  2.611e+07   0.546 0.585211
TeamSan Diego Padres      6.816e+06  2.385e+07   0.286 0.775279
TeamSan Francisco Giants -3.338e+07  2.032e+07  -1.643 0.101489
TeamSeattle Mariners      2.142e+07  2.213e+07   0.968 0.333895
TeamSt. Louis Cardinals   5.782e+07  2.546e+07   2.271 0.023908 *
TeamTampa Bay Rays        1.775e+07  2.742e+07   0.647 0.518028
TeamTexas Rangers         2.079e+07  1.759e+07   1.182 0.238230
TeamWashington Nationals -5.414e+07  3.196e+07  -1.694 0.091377 .
Year1999                  3.974e+06  8.465e+06   0.469 0.639117
Year2000                 -6.302e+06  8.736e+06  -0.721 0.471281
Year2001                 -7.414e+06  8.693e+06  -0.853 0.394469
Year2002                 -1.092e+07  8.904e+06  -1.226 0.221218
Year2003                 -1.679e+07  9.284e+06  -1.808 0.071640 .
Year2004                 -1.794e+07  1.042e+07  -1.721 0.086308 .
Year2005                 -2.157e+07  1.175e+07  -1.836 0.067398 .
Year2006                 -3.021e+07  1.394e+07  -2.168 0.030993 *
Year2007                 -2.826e+07  1.549e+07  -1.824 0.069219 .
Year2008                 -2.795e+07  1.635e+07  -1.709 0.088453 .
```

```
Year2009                   -2.216e+07  1.436e+07  -1.543 0.123953
Year2010                   -3.151e+07  1.512e+07  -2.084 0.038079 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23270000 on 289 degrees of freedom
Multiple R-squared:  0.8706,     Adjusted R-squared:  0.8415
F-statistic: 29.91 on 65 and 289 DF,  p-value: < 2.2e-16
```



Appendix 2: In pursuit of statistical significance for our regression coefficients.

Introduction

The primary concern with the results obtained so far is the lack of significance in the coefficients of interest, resulting in lack of meaningful insights as a whole. In this section, we will try to overcome this problem by exploring limiting the dataset in a way that leads to more significant results with our IV regression, and exploring other methodologies that sufficiently account for endogeneity without using the IV.

In this section we will try to identify means for dealing with the weak IV problem, by possibly limiting the scope and applicability of our results. In first part of this section, we will try to identify weaknesses in our IV's ability to predict Wins in the first stage regression, and explore if accounting for heterogeneity in characteristics of teams makes the IV stronger. In the second part of this section, we explore an alternative methodology for obtaining our lagged coefficients that directly avoids the reverse causality and endogeneity concerns.

Improving IV strength

One of the curious observations with the opposition strength IV is that there isn't clear negative relationship between opposition strength and wins in a given season for teams with high wins. In particular we note that teams with wins between 85 and 95, the relationship become strongly positive, which is very hard to explain. Whereas for teams with wins less than 85, the opposition strength IV has a strongly positive relationship and a higher R^2.
Results:

	75>W>80	80>W>85	85>W>90	90>W>95	100>W>95
Correlation	-0.05	-0.08	-0.09	0.24	-0.23
beta	-0.45	-0.66	-0.78	2.37	-1.66
T-value	-0.54	-0.751	-0.99	2.36	-1.99
R^2	.0028	.0063	0.008	0.058	0.056

F-value	.2943	0.564	0.984	5.61	3.64

```
lm(formula = reg_data[Year >= 1996 & Year <= 2016]$W ~ reg_data[Year >=
    1996 & Year <= 2016]$strength_of_schedule)

Residuals:
    Min      1Q  Median      3Q     Max
-38.144  -8.697  -0.035   8.450  34.856

Coefficients:
                                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                                         81.1438     0.4575 177.351  < 2e-16 ***
reg_data[Year >= 1996 & Year <= 2016]$strength_of_schedule  -7.0293     2.3087  -3.045  0.00243 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.25 on 603 degrees of freedom
Multiple R-squared:  0.01514,	Adjusted R-squared:  0.01351
F-statistic:  9.27 on 1 and 603 DF,  p-value: 0.002431
```

One hypothesis for this observation is that for this cohort of teams, which is very close to the playoff threshold, higher opposition strength also results in higher threshold for winning the division, and a higher threshold in season with higher opposition strength results in teams striving for and achieving a higher number in spite of the greater opposition strength (possibly because the higher threshold incentivizes them to win more of the close clutch games).

This leads us to controlling for Division_win_threshold in the first stage IV regression, where Division_win_threshold is defined as the number of wins by the Division leader at the end of the season.

We observe, that adding this control does in fact improve the F-stat of our first stage regression, and we no longer have the anomalous positive relationship between opposition strength and wins for teams closer to the playoff cutoff threshold.

	Dependent variable:	
	W	
	(1)	(2)
strength_of_schedule	-7.029***	-9.922***
	(2.309)	(2.400)
division_win_thresh		0.416***
		(0.107)
Constant	81.144***	42.461***
	(0.458)	(9.945)
Observations	605	605
R2	0.015	0.039
Adjusted R2	0.014	0.036
Residual Std. Error	11.252 (df = 603)	11.122 (df = 602)
F Statistic	9.270*** (df = 1; 603)	12.325*** (df = 2; 602)
Note:		*p<0.1; **p<0.05; ***p<0.01

The new first stage regression has an F score of 12.32 compared the previous score of 9.27

	Dependent variable:	
	W	
	(1)	(2)
strength_of_schedule	2.371**	0.834
	(1.001)	(0.890)
division_win_thresh		0.367***
		(0.062)
Constant	92.518***	58.393***
	(0.186)	(5.749)
Observations	92	92
R2	0.059	0.326
Adjusted R2	0.048	0.311
Residual Std. Error	1.779 (df = 90)	1.514 (df = 89)
F Statistic	5.610** (df = 1; 90)	21.506*** (df = 2; 89)
Note:		*p<0.1; **p<0.05; ***p<0.01

Above regression is looking only at the dataset where 95>W>100, which had anomalous results previously. We see that adding the division_win_threshold control causes the anomalous positive and significant beta to now be statistically indistinguishable from 0. Also we see huge jump in R^2 and F values.

Another, interesting observation is that Division_win_threshold can also be used as a second IV, since Division_win_threshold should not related to Revenue through any mechanism other than its affect on wins.

Having a second IV, also makes it possible to include both playoffs and wins as our endogenous variables in the regression, and therefore be able to directly disentangle the effect of wins from playoffs on next year's revenues.

Results:

Original 1IV regression with and without controls below:

	Dependent variable:	
	real_rev_year_plus_1	
	(1)	(2)
W	4,144,851.000**	5,673,743.000***
	(1,670,783.000)	(1,599,433.000)
population		1.215
		(0.976)
per_capita_income		1,253.703***
		(472.560)
stadium_age		378,805.500***
		(118,949.300)
Year		480,717.200
		(775,390.000)
Constant	-208,233,435.000	-1,365,729,679.000
	(135,604,465.000)	(1,590,706,876.000)
Observations	576	576
R2	0.075	-0.107
Adjusted R2	0.074	-0.117
Residual Std. Error	56,026,415.000 (df = 574)	61,517,120.000 (df = 570)
Note:		*p<0.1; **p<0.05; ***p<0.01

Same regression but with 2IV's below:

	Dependent variable:	
	real_rev_year_plus_1	
	(1)	(2)
W	2,526,700.000***	3,275,470.000***
	(919,517.200)	(845,179.100)
population		2.381***
		(0.607)
per_capita_income		1,541.889***
		(342.065)
stadium_age		460,887.200***
		(84,977.420)
Year		42,497.430
		(564,926.000)
Constant	-76,920,125.000	-313,060,175.000
	(74,650,514.000)	(1,141,941,533.000)
Observations	576	576
R2	0.202	0.366
Adjusted R2	0.201	0.361
Residual Std. Error	52,048,555.000 (df = 574)	46,545,036.000 (df = 570)
Note:		*p<0.1; **p<0.05; ***p<0.01

2IV regression, with Wins and Post_season_appearance as endogenous variables below:

```
Call:
ivreg(formula = real_rev_year_plus_1 ~ post_season_appearance +
    W + per_capita_income + stadium_age + Year | . - post_season_appearance -
    W + strength_of_schedule + division_win_thresh + wild_card_thresh,
    data = reg_data[Year >= 1996 & Year <= 2016])

Residuals:
      Min         1Q     Median         3Q        Max
-213900400  -30254700    3375555   36908949  273692963

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.136e+09  1.642e+09   0.692 0.489493
post_season_appearance 1.262e+08  6.765e+07   1.865 0.062678 .
W                      8.774e+05  1.794e+06   0.489 0.625029
per_capita_income      2.206e+03  5.054e+02   4.365 1.51e-05 ***
stadium_age            4.452e+05  1.303e+05   3.416 0.000682 ***
Year                  -6.070e+05  8.002e+05  -0.759 0.448447
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 63350000 on 570 degrees of freedom
Multiple R-Squared: -0.1738,    Adjusted R-squared: -0.1841
Wald test: 29.01 on 5 and 570 DF,  p-value: < 2.2e-16
```

We note that because of the additional IVs we now have identification for both Wins and post_season_appearance endogenous factors, however we lose significance for both of them. The reason for this is that wins and post season appearance are very highly correlated, and cause multicollinearity issue. (Clearly high wins, also increases the probability of going to the playoffs).

We note that we no longer have statistical significance for either of our variables of interest (wins and playoffs), because of the strong correlation between wins and playoff appearance, since clearly high number of wins is correlated with playoff appearance.

In pursuit of finding significance in the regression coefficients, so that we may be able to say something statistically meaningful, we will look at only a subset of data, with the hope that there is more signal than noise when we narrow our vision.

Next we observe, that if we restrict the IV to the subset of teams with a last4 season average wins between 65 and 90, we are able to extract more signal and get much higher statistical significance for wins and playoffs.

This allows us to say something useful about what the impact of higher wins and playoff appearance will be on next year's revenues for teams average historical wins in the range of 65-90. Since majority of the teams are in this category anyways, this essentially helps us take out outliers that are potentially adding more noise than signal.

```
Call:
ivreg(formula = real_rev_year_plus_1 ~ post_season_appearance +
    W + per_capita_income + stadium_age | . - post_season_appearance -
    W + strength_of_schedule + division_win_thresh + wild_card_thresh,
    data = reg_data[Year >= 1996 & Year <= 2016][avg_wins_past4 >=
        65 & avg_wins_past4 <= 90])

Residuals:
      Min         1Q     Median         3Q        Max
-206770776  -32525140    8023363   39877244  125497712

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -2.379e+08  1.242e+08  -1.916   0.0560 .
post_season_appearance  9.056e+07  5.147e+07   1.759   0.0792 .
W                       3.437e+06  1.674e+06   2.053   0.0407 *
per_capita_income       1.291e+03  3.101e+02   4.163 3.75e-05 ***
stadium_age             5.357e+05  1.269e+05   4.222 2.92e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 62790000 on 464 degrees of freedom
Multiple R-Squared: -0.5858,    Adjusted R-squared: -0.5995
Wald test: 20.45 on 4 and 464 DF,  p-value: 1.551e-15
```

Here we note that the coefficient for Wins is once again significant even with the addition of highly correlated post_season_appearance variable. Moreover the post_season_appearance coefficient is also significant at the 8% level, which is reasonable level of confidence.

The value of the Wins coefficient seems pretty robust and holds up to the addition of additional IVs and additional covariates at ~ 3.5M, indicating that each additional win brings about ~3.5M in additiona regular season revenue next year.

The value of the post_season_appearance coefficient is about $90M, indicating that going to the playoff brings additional $90M of regular season revenue net year.

Alternative methodology:

We observe that our primary challenge is accounting for reverse causality that manifests itself in the relationship between richer metropolitan areas and higher wins, which operates through the mechanism of higher revenues for riches metropolitan areas allows teams to get better players.

We note that this relationship strongly holds for the levels, but is not important for first differences in time. Teams which see an increase in their level of wins (compared to last few season's benchmark), can be expected to deliver higher revenues in the next season, but it is reasonable to postulate that an increase revenues next season can causally result in increase in wins this season. One possible counter-situation is if a team is on a sustained growth path (in terms of the number of wins), and therefore future growth in revenues can be anticipated by management resulting in them getting acquiring better and better players every year. This is an unsustainable situation, and as long as teams reach a steady state in the number of wins, it can be argued that unexpected increase in the number of wins will causally increase revenues next season, but in steady state the increase in revenues is unexpected.

A simple test of this hypothesis is to check if there is an external variable that is related to both a high delta in revenues and a high delta in wins. Such a variable will then result in endogeneity, if not properly controlled for. We were not able to think of any such likely variable, primarily because of the fact that any shock to wins will eventually level off and result in sustain continued growth in wins and therefore won't result in continued positive delta_wins.

In this regression we are looking at deltas and not levels, and therefore we need to identify a benchmark to calculate the deltas against. We use the average of last 3 years' revenues as the revenue benchmark, and the last 3 years' of wins as wins benchmark. We define delta_wins and delta_revenues as the difference between the level values in our year of interest and the lagging benchmark as defined above. We also define PS_delta as difference between reaching the post-season and expectation of reaching the playoff based on last 3 years track record of reaching the playoffs.

Our hypothesis is that this methodology does not suffer from endogeneity issues and we will test that by adding covariates, and observing the significance of their coefficients and the effect on our coefficients of interest.

We also run two different forms of regressions:

The first regression explores the effect of delta_wins today on delta_revenues today, and delta_revenues upto 5 years in the future. In this regression we are not controlling for performance in the future years, and therefore the coefficient obtained should be considered to account for all channels through which higher wins today manifest themselves (such as network effect attracting better players, persistence of increased fan loyalty, and team momentum carrying over and resulting in greater wins in following years.)

	Dependent variable:			
	(Revs_delta)			
	(1)	(2)	(3)	(4)
Wins_delta	711,636.000***	708,261.900***	700,949.900***	686,640.700***
	(124,931.900)	(124,483.500)	(125,340.500)	(124,514.700)
PS_delta	-3,503,997.000	-3,357,798.000	-3,281,541.000	-3,010,172.000
	(3,019,035.000)	(3,008,682.000)	(3,013,432.000)	(2,992,758.000)
WS_winner	19,984,364.000***	18,628,584.000***	19,346,683.000***	17,683,927.000***
	(6,040,388.000)	(6,048,188.000)	(6,047,245.000)	(6,028,294.000)
population		0.513**	0.645***	0.463*
		(0.227)	(0.233)	(0.239)
per_capita_income			-297.597***	-297.837***
			(110.741)	(109.931)
stadium_age				129,099.500***
				(42,796.140)
Constant	4,234,834.000***	1,342,999.000	13,821,970.000***	11,583,578.000**
	(1,077,861.000)	(1,671,922.000)	(4,871,581.000)	(4,892,568.000)
Observations	560	560	554	554
R2	0.094	0.102	0.112	0.127
Adjusted R2	0.089	0.095	0.104	0.117
Residual Std. Error	25,036,470.000 (df = 556)	24,944,824.000 (df = 555)	24,906,022.000 (df = 548)	24,723,970.000 (df = 547)
F Statistic	19.135*** (df = 3; 556)	15.730*** (df = 4; 555)	13.827*** (df = 5; 548)	13.210*** (df = 6; 547)
Note:				*p<0.1; **p<0.05; ***p<0.01

Revs_delta(t) ~ Wins_delta(t) + PS_delta(t) + WS_win(t) + controls

In the above series of regressions, we are trying to check if any of the primary control candidates are correlated with our endogenous variable wins_delta, PS_delta, and WS win. It is interesting to note that the coefficient for wins_delta is robust to the inclusion of controls such as population, per-capita-income and stadium_age, indicating that these variables are orthogonal wins_delta, and will not lead to endogeneity. WS_win and PS_delta are also holding steady, and therefore we do not see a need to control for them in subsequent regressions.

Next we look at 1 through 5 year lagged effect on revenues, i.e. we ask the question what will be delta_revenues 1-5 years from now compared with the current year's benchmark, and not controlling for future performance shocks.

	Dependent variable:					
	(Revs_delta)	(Revs_delta_year_plus_1)	(Revs_delta_year_plus_2)	(Revs_delta_year_plus_3)	(Revs_delta_year_plus_4)	(Revs_delta_year_plus_5)
	(1)	(2)	(3)	(4)	(5)	(6)
Wins_delta	711,636.000***	1,195,980.000***	1,207,184.000***	1,069,474.000***	818,430.700***	874,335.700***
	(124,931.900)	(159,721.900)	(197,568.200)	(229,082.900)	(260,036.300)	(289,639.000)
PS_delta	-3,503,997.000	-755,039.700	3,136,495.000	3,708,742.000	1,401,820.000	-2,216,763.000
	(3,019,035.000)	(3,882,556.000)	(4,800,773.000)	(5,547,920.000)	(6,321,849.000)	(7,030,874.000)
WS_winner	19,984,364.000***	18,515,164.000**	16,941,409.000*	14,816,747.000	20,270,534.000	20,023,665.000
	(6,040,388.000)	(7,726,921.000)	(9,524,101.000)	(11,031,940.000)	(12,443,639.000)	(13,776,560.000)
Constant	4,234,834.000***	6,397,704.000***	8,816,505.000***	10,763,511.000***	12,452,583.000***	15,211,847.000***
	(1,077,861.000)	(1,378,928.000)	(1,699,432.000)	(1,971,130.000)	(2,229,665.000)	(2,471,155.000)
Observations	560	530	500	470	440	410
R2	0.094	0.160	0.133	0.090	0.048	0.038
Adjusted R2	0.089	0.155	0.127	0.084	0.041	0.031
Residual Std. Error	25,036,470.000 (df = 556)	31,157,202.000 (df = 526)	37,293,781.000 (df = 496)	41,938,364.000 (df = 466)	45,901,263.000 (df = 436)	49,105,980.000 (df = 406)
F Statistic	19.135*** (df = 3; 556)	33.448*** (df = 3; 526)	25.306*** (df = 3; 496)	15.389*** (df = 3; 466)	7.281*** (df = 3; 436)	5.319*** (df = 3; 406)
Note:						*p<0.1; **p<0.05; ***p<0.01

We get very promising results here, particularly for the coefficient associated with Wins_delta which is strongly significant: in particular we note that exceptional performance today leads to slightly higher revenue this year, but much improved revenue in the subsequent year when the fan excitement has built up. Morevover the increase in revenue decreases linearly over the next 5 years, as expected.

We note that PS_delta is not statistically significant, hence cannot be differentiated from 0. This is a result of the collinearily betwee wins_delta and PS_delta: a big wins_delta and increases the chances of an associated PS_delta.

To get more signal out of PS_delta, we would like to take advantange of heterogeneity between the poorly performing and well performing teams in the impact making the playoffs has on the revenues. In particular, we hypothesize that poorly performing teams will have more of a boost from making the playoffs than teams that had 50/50 chance of making the playoffs.

We define wins_gap as the distance between the threshold for making the playoffs and the teams benchmark from previous performance. We then add this as an interaction term to PS_delta term. Finally we restrict our data to teams with wins <90, to isolate this effect which would be expected to reverse for well performing teams, and thus confound our coefficient of interest (unless we include quadratic terms)

	Dependent variable:					
	(Revs_delta) (1)	(Revs_delta_year_plus_1) (2)	(Revs_delta_year_plus_2) (3)	(Revs_delta_year_plus_3) (4)	(Revs_delta_year_plus_4) (5)	(Revs_delta_year_plus_5) (6)
Wins_delta	648,995.100***	1,164,198.000***	1,242,959.000***	1,237,451.000***	1,014,627.000***	1,033,388.000***
	(134,873.300)	(177,400.100)	(215,456.700)	(254,352.300)	(282,294.000)	(306,746.000)
PS_delta	-11,366,555.000***	-9,414,516.000*	1,258,591.000	4,353,578.000	131,738.400	-6,156,512.000
	(3,735,820.000)	(4,944,403.000)	(6,067,005.000)	(7,034,572.000)	(7,775,568.000)	(8,475,746.000)
WS_winner	23,158,930.000*	26,766,216.000*	22,172,209.000	24,061,191.000	29,297,110.000	28,923,117.000
	(12,412,392.000)	(15,909,703.000)	(18,750,224.000)	(24,439,385.000)	(26,319,617.000)	(27,660,163.000)
PS_delta:Wins_gap	988,005.700***	1,420,127.000***	611,831.700	290,001.300	-136,884.200	-106,628.000
	(379,365.800)	(495,251.400)	(655,562.400)	(761,135.900)	(835,129.400)	(899,665.000)
Constant	843,406.200	3,291,035.000*	8,377,447.000***	11,896,462.000***	13,511,297.000***	15,402,155.000***
	(1,282,523.000)	(1,689,985.000)	(2,046,196.000)	(2,398,938.000)	(2,644,458.000)	(2,867,757.000)
Observations	430	406	382	356	336	314
R2	0.068	0.123	0.109	0.089	0.049	0.040
Adjusted R2	0.059	0.114	0.099	0.079	0.038	0.027
Residual Std. Error	24,257,055.000 (df = 425)	31,045,233.000 (df = 401)	36,522,881.000 (df = 377)	41,451,975.000 (df = 351)	44,586,524.000 (df = 331)	46,764,734.000 (df = 309)
F Statistic	7.745*** (df = 4; 425)	14.002*** (df = 4; 401)	11.520*** (df = 4; 377)	8.623*** (df = 4; 351)	4.303*** (df = 4; 331)	3.192** (df = 4; 309)
Note:						*p<0.1; **p<0.05; ***p<0.01

Here we note that we have managed to extract signal from the PS_delta variable. Although the PS_delta term is negative, we also need to add the interaction coefficient multiplied by Wins_delta. The negative results for PS_delta are troubling however, and we do not know how to interpret this properly.

These coefficient is of interest for a GM who is trying to looking for a way to predict aggregate increase in revenues upto 5 years in the future.

Our second regression explores the effect of delta_wins today on delta revenues upto 5 years in the future, but controlling for performance in the intervening years. The interpretation of this regression is that we are accounting for channels such as getting better players or continued team momentum, which results in better performance and more wins than our benchmark in the subsequent years. The coefficient of our regression then can be interpreted as having isolated the effect of just the increased fan motivation over the next 5 years, which does not manifest itself in better performance (assuming here that home field advantage related effect is minimal).

	Dependent variable:			
	(Revs_delta_year_plus_1)	Revs_delta_year_plus_1	(Revs_delta_year_plus_2)	(Revs_delta_year_plus_3)
	(1)	(2)	(3)	(4)
Wins_delta	1,195,980.000***	955,580.400***	608,078.300***	332,305.500
	(159,721.900)	(180,088.000)	(216,272.800)	(243,909.000)
PS_delta	-755,039.700	-2,513,381.000	-17,245.660	-532,251.800
	(3,882,556.000)	(4,010,141.000)	(4,804,918.000)	(5,424,924.000)
Wins_delta_year_plus_1		397,997.800**	666,512.200***	408,537.200
		(167,668.300)	(219,226.500)	(249,160.100)
PS_delta_year_plus_1		2,935,616.000	5,247,312.000	7,862,745.000
		(3,811,970.000)	(4,666,218.000)	(5,289,128.000)
Wins_delta_year_plus_2			291,484.100	693,301.900***
			(202,067.900)	(229,257.100)
PS_delta_year_plus_2			2,561,654.000	7,090,706.000
			(4,622,277.000)	(5,297,508.000)
WS_winner	18,515,164.000**	21,244,967.000***	21,770,019.000**	19,331,492.000*
	(7,726,921.000)	(7,687,232.000)	(9,116,419.000)	(10,334,100.000)
Constant	6,397,704.000***	6,336,755.000***	8,500,301.000***	10,259,986.000***
	(1,378,928.000)	(1,366,338.000)	(1,620,725.000)	(1,839,973.000)
Observations	530	529	498	468
R2	0.160	0.182	0.213	0.206
Adjusted R2	0.155	0.174	0.201	0.194
Residual Std. Error	31,157,202.000 (df = 526)	30,834,851.000 (df = 523)	35,475,456.000 (df = 490)	39,040,868.000 (df = 460)
F Statistic	33.448*** (df = 3; 526)	23.298*** (df = 5; 523)	18.908*** (df = 7; 490)	17.100*** (df = 7; 460)
Note:				*p<0.1; **p<0.05; ***p<0.01

The coefficients of interest in this regression are the ones for Wins_delta. We can see that their magnitude are decreasing and lower than the ones in the first set of regressions we performed, where we were not controlling for future performance. This indicates that the effect of the fans only channel is lower than the combined effect of all the channels on future revenues, which is what we would logically expect.

This coefficient of interest is for a GM, who considers the improvement in performance this year to not be sustainable (or a fluke), and thus only wants to account for increased revenues stemming only from greater fan motivation.

Or primary concerns with this regression is our calculation of the benchmark to measure delta wins and revenues against. We are taking average of last 3 years of performance, and to the extent that there were abnormal shocks in the previous years, this benchmark will be biased upwards. Several assumptions are needed to justify this choice of benchmark: 3 years is a sufficiently large time period to have players and fan's expectations set in, and therefore is a fair benchmark for measuring how far above or below expectations the team performed; we also need to assume that economy's growth is constant over the 3 years period and there aren't (on average) increase (decrease) in the level of growth rates, which would imply accelerating (decelerating) revenues, which would again bias our benchmark.

In terms of possible source of endogeneity causing omitted variable bias, we believe our primary concern is that emerging/ new teams that are trying to establish themselves might have long sustained, high improvements in performance and also increases in revenue which have a life of their own and not necessarily related to increases in performance, and thus would upward bias our results. We ignored Washington Nationals in these regressions, but this can also apply to teams in economically declining cities like Detroit, where revenues are decreasing and good players are not interested in living in the city resulting in a correlation between shocks to revenues and winning, but not a causal connection between the two other. (Again this would upward bias our measurement). However our team fixed effects analysis indicates that team/ brand equity is relatively constant and persistent over time, and therefore we should not need to control for these type of scenarios.

Conclusion:

In conclusion, this appendix attempts looked more narrowly at subset of the data in an attempt to obtain significance for our IV regression variables. We accomplish that for teams with average win between 65 and 90 per season, getting additional revenue from additional win around ~3.5M, and post_season_appearance ~90M. However these values have T-ratois very close to 2, and hence there is a significant amount of uncertainty associated with these values. However, we can say that they are more than 0.

The alternative methodology seeks to get an even greater level of precision be looking at first differences rather than levels. We get much higher precision and and confidently say that each additional win will yield between 600k and 1.400M in additional revenue. This seems farther away from ~3.5M obtained via 2IV regression, but if we account for the large std errors of the IV regression, these results are not inconsistent.