The Implications of Tax Policy on NBA Franchise Success

Paper by Timothy Zimmer

Write Up and Replication by Preston Klaus

SAL 366: Dr. Jeremy Losak

Introduction:

In this paper, we examine the relationship between income tax rates and the success of National Basketball Association (NBA) franchises. In the last two Collective Bargaining Negotiations, there has been turmoil between large and small market teams about leveling the playing field for smaller market organizations. In 2011, it centered around the luxury tax, as smaller market teams had fewer resources to exceed the tax. In 2016, teams were now able to extend current players on their roster earlier and for longer to prevent them from leaving for bigger markets in free agency. However, there may be a hidden variable: the tax brackets of each major city. The Collective Bargaining Agreement (CBA) is put in place to try and level the playing field for every team with a salary cap, giving small and large market teams the same number of resources. Thus, if higher tax brackets are a disadvantage to certain teams, then the CBA is not fulfilling its purpose and needs to be reexamined.

In the original paper, Timothy Zimmer was able to conclude that tax brackets have a strong indirect negative influence on winning percentage. His binary year variables were all insignificant, ruling out the existence of significant yearly variation. In addition, Zimmer found a significant negative correlation between the two years lagged winning percentage, suggesting league parity, and the coaching change variable, suggesting a coaching change is negative in the short term. The draft pick variable was insignificant, which I believe is due to teams not getting fair and equal value for their pick in a draft pick transaction, going against what Zimmer assumed. Overall, Zimmer argued that the CBA should include a tax adjustment index in order to correct for this disparity and further research needed to be done in order to confirm this result.

Ten years have passed since this original study was released and the league has changed drastically. Every team has embraced the use of analytics to make decisions, two new CBAs have been approved, one resulting in a lockout, in the last decade, and pre-coronavirus league revenues were more than double that of the ones referenced in the paper in 2011, largely due to a huge TV deal signed before the 2015-2016 season. These changes have altered league decision making, salary structures, and, most glaringly, pay. In my paper, I decided to not only add two more variables, but also use newer data in order to test this same hypothesis. I believe this is critical because the average NBA salary has increased drastically over recent years. More importantly, the NBA minimum salary has grown tremendously. The minimum salary increases on a scale based on a player’s years of experience in the NBA, from 0 to 10+. In twenty years, the minimum salary has jumped from $316,969 to $898,310 for a player with no years of experience. Additionally, a player with one year of experience is making $1,445,697 today, compared to a player with ten years of experience making a meager $1,000,000, in comparison, in 2001. Therefore, in Zimmer’s original study, using the top marginal tax bracket for each state is slightly flawed because many players did not fall under this tax bracket. By using updated data, almost all NBA players fall under this maximum tax bracket for each respective state.

Methodology:

In total, there were four models created for this paper. Two models were run for every NBA team, excluding the ones from Canada, from the years 2001 to 2010, like that in the original study with the same variables. However, one predicted the actual winning percentage of each team and the other predicted what Zimmer refers to as “Z”, which rounds win percentage to an integer. Zimmer provides results for the former but references both equations within his paper. Thus, I decided to run both models.

$$\text{Win Pct}_{i,t} = \alpha + \beta_1(\text{Year}_t) + \beta_2(\text{Win Pct}_{i,(t-1)}) + \beta_3(\text{Win Pct}_{i,(t-2)}) + \beta_4(\text{Win Pct}_{i,(t-3)}) + \beta_5(\text{Coach}_{i,t}) + \beta_6(\text{Coach}_{i,(t-1)}) + \beta_7(\text{DraftSqr}_{i,t}) + \beta_8(\text{DraftSqr}_{i,(t-1)}) + \beta_9(\text{Metro Size}_{i,t}) + \beta_{10}(\text{StateIncTax}_{i,(t-1)}) + \varepsilon$$

$$Z = \alpha + \beta_1(Year_t) + \beta_2(Win\ Pct_{i,(t-1)}) + \beta_3(Win\ Pct_{i,(t-2)}) + \beta_4(Win\ Pct_{i,(t-3)}) + \beta_5(Coach_{i,t}) + \beta_6(Coach_{i,(t-1)}) + \beta_7(DraftSqr_{i,t}) + \beta_8(DraftSqr_{i,(t-1)}) + \beta_9(Metro\ Size_{i,t}) + \beta_{10}(StateIncTax_{i,(t-1)}) + \varepsilon$$

The other two models similarly predict winning percentage and the variable "Z", defined earlier, but for the years between 2010-2019. Additionally, I included two new variables into the models: Pre-Draft Trade for a First Round Draft Pick and Growth Percentage of The Metro Area.

$$Win\ Pct_{i,t} = \alpha + \beta_1(Year_t) + \beta_2(Win\ Pct_{i,(t-1)}) + \beta_3(Win\ Pct_{i,(t-2)}) + \beta_4(Win\ Pct_{i,(t-3)}) + \beta_5(Coach_{i,t}) + \beta_6(Coach_{i,(t-1)}) + \beta_7(DraftSqr_{i,t}) + \beta_8(DraftSqr_{i,(t-1)}) + \beta_9(Metro\ Size_{i,t}) + \beta_{10}(StateIncTax_{i,(t-1)}) + \beta_{11}(PreDraftTrd_{i,t}) + \beta_{12}(MetroGrwth_{i,t}) + \varepsilon$$

$$Z = \alpha + \beta_1(Year_t) + \beta_2(Win\ Pct_{i,(t-1)}) + \beta_3(Win\ Pct_{i,(t-2)}) + \beta_4(Win\ Pct_{i,(t-3)}) + \beta_5(Coach_{i,t}) + \beta_6(Coach_{i,(t-1)}) + \beta_7(DraftSqr_{i,t}) + \beta_8(DraftSqr_{i,(t-1)}) + \beta_9(Metro\ Size_{i,t}) + \beta_{10}(StateIncTax_{i,(t-1)}) + \beta_{11}(PreDraftTrd_{i,t}) + \beta_{12}(MetroGrwth_{i,t}) + \varepsilon$$

Win Percentage has a discrete distribution, so the variable Z corrects for that by assuming Win Percentage is "a manifestation of an underlying continuous variable," according to Zimmer. In simpler terms, it is displayed as a continuous variable. Since this paper is working with a specific subset of NBA data that can't be applied to other leagues, this edit is Zimmer's way of capturing the fixed effects of the model and the method I will also be adopting with later seasons.

Year is a binary dummy fixed effects variable that references the timeframe the data is from. The year 2005 is dropped from Zimmer's model, and the years 2001 and 2010 are dropped in my first two and last two models, respectfully, to prevent linear dependency.

Winning Percentage was lagged by one, two, and three years for each "i" team and acquired off of landofbasketball.com. Since the NBA season spans over two years, the last year was used as a lag term. For example, for the 2011-2012 season, the predicted win percentage dependent variable would fall under the year 2012 and the win percentage lagged one year would be that team's record from the 2010-2011 season.

Coaching Change is a binary variable with “1” signifying a coaching change in the offseason or regular season of that given year and “0” signifying no coaching change. This is in the model as a regular and lagged term. If there was a coaching change within the offseason and regular season for one team in a given year, I still labeled this variable as “1”, even though there were two coaching changes. To account for a one-year coaching adjustment, the lagged coaching change variable is included. This data was compiled off of wikipedia.com.

Draft Weight is listed in the model as DraftSqr as a regular and lagged term. This data was compiled off of prosportstransactions.com based on the original draft pick slot for each team, before trades. The draft information was taken from the draft the offseason before, as the rookies play their first minutes in the subsequent season. For example, in the 2011-12 season, the draft data was compiled from the 2011 draft. The lagged draft data would come from 2010 and accounts for one year of that rookie’s production, allowing top prospects to mature and grow after one year on a team. The second round is not included because contracts are non-guaranteed from that round and first round picks are on each team for at least two seasons. Then, each pick is assigned inverse weights for each season per team, with pick 1 having a value of 30, 2 having a value of 29, etc. assuming there are thirty first round picks. These values are then squared to add emphasis to the earlier picks within the draft.

The Metro Size variable was compiled from macrotrends.net and accounts for bias in larger markets. A New York market may be more or less favorable for a player compared to that of Oklahoma City because of the population size.

The top marginal state income tax rate was compiled off of taxfoundation.org and is lagged one year because players will sign contracts based on data from the year before. It will measure taxes on overall team performance.

The first variable added was a binary Pre-Draft Trade variable, which was acquired manually based on data from prosportstransactions.com. “1” signifies that the draft pick was traded before the draft started, while “0” signifies the draft pick was not traded.

The second variable added was one for lagged growth percentage of the metro size. This variable was acquired from macrotrends.net and is lagged because players will sign contracts based on trends from the year before.

There were multiple critiques I had with the original paper. First, 2001 was the last season the Grizzlies were in Vancouver, and the paper gave no reasoning behind what they did for observation. I ended up excluding the value because the team was in Canada at the time but was unsure what to do about the lagged tax bracket data and the lagged win percentages from when the team was still in Canada.

From 2001-2004, there were only 29 draft picks, so weights were assigned from 29 to 1 and then squared. Once the Charlotte Bobcats entered the league in 2004, the non-squared weight given to the number one pick was 30. This observation was left out because there were no years to lag before it, the draft pick was given not based off of record and it would have skewed the data. The Minnesota Timberwolves also forfeited draft picks in 2001, 2002, and 2004 because of a violation. For this case, the original paper didn’t specify what to do, so I gave that pick (or lack of one) a value of 0.

For draft picks, the methodology is slightly flawed for each draft pick weight. Draft picks don’t necessarily follow an exponential distribution and some pick selections are determined through the lottery or through tiebreakers. These changes would be tough to definitively correct for. The adjusted R-squared value was not listed, but the R-squared value of this paper wasn’t very high. While there are limitations to adjusted R-squared, not having a high mark is a concern for this model.

I ended up correcting three limitations in the model. First, I included a binary pre-draft trade variable to account for a trade before the draft because trades are inherently unequal. This notably does not include trades that occurred on draft night because I believe the effects are accounted for in other variables. Next, I included a variable for lagged Metro Growth because players may be inclined to go to growing cities, rather than declining cities. This variable is lagged because a player would make a decision to potentially sign based on data from the year before. If that player thinks the city is on a positive trend, they may be more inclined to sign. I also thought about including the lagged crime rate for each city but didn’t feel it would have much impact and that it may be accounted for in this new variable. Lastly, I ran regressions with updated years because of exponentially increasing contract and cap figures. Because some tax brackets in states like California, New York, and New Jersey have the highest tax bracket for residents earning upwards of $1,000,000, it was flawed to assume each player’s decisions based on that top tax bracket number. Collectively, all minimum contract players from 2001 who were playing on the six teams within those three states, were not paying that tax level. Additionally, rookie contract players making $316,969 in 2001 would not fall under a top tax bracket of income over $500,000. Because this salary number was so far from the top tax bracket number for so many players, I felt it was necessary to run this model with current observations. These years would also capture the effects of two lockouts and the analytics movement within the NBA.

Results for the Replication:

Zimmer posted his winning percentage estimates in his paper, including negligible Year fixed effects in the regression results table. I excluded my year variable in the replication regression because it was insignificant.

	Dependent variable:	
	Winning Percentage	Z Model
	OLS	*logistic*
	(1)	(2)
Winning Percentage (lag 1)	0.384^{***}	4.425^{**}
	(0.076)	(1.785)
Winning Percentage (lag 2)	-0.023	-0.790
	(0.054)	(1.261)
Winning Percentage (lag 3)	0.057	1.815
	(0.051)	(1.125)
Coaching Change	-0.101^{***}	-1.941^{***}
	(0.016)	(0.364)
Coaching Change (lag 1)	-0.018	-0.553
	(0.016)	(0.371)
Draft Pick Weight Squared	-0.00004	-0.002^{**}
	(0.00004)	(0.001)
Draft Pick Weight Squared (lag 1)	0.00001	0.0004
	(0.00003)	(0.001)
Market Size	-0.001	-0.025
	(0.001)	(0.029)
Tax Bracket (lag 1)	-0.483^{**}	-10.008^{*}
	(0.219)	(5.342)
Constant	0.369^{***}	-0.169
	(0.066)	(1.516)
Observations	284	284
R^2	0.450	
Adjusted R^2	0.413	
Log Likelihood		-128.023
Akaike Inf. Crit.		294.047
Residual Std. Error	0.114 (df = 265)	
F Statistic	12.049^{***} (df = 18; 265)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 3: Basketball Team Winning Percentage Estimation

Variable	Coefficient	Standard Error
Constant	0.2464	
2001	-0.0175	0.0310
2002	-0.0162	0.0309
2003	-0.0290	0.0309
2004	-0.0050	0.0310
2006	-0.0105	0.0303
2007	-0.0379	0.0306
2008	-0.0279	0.0308
2009	-0.0094	0.0304
2010	-0.0195	0.0300
Win Pct Lagged 1 Year ($WinPct_{i,(t-1)}$)	0.5656	0.0942 ***
Win Pct Lagged 2 Year ($WinPct_{i,(t-2)}$)	0.1586	0.0963 *
Win Pct Lagged 3 Year ($WinPct_{i,(t-3)}$)	-0.1148	0.0587 **
Coach ($Coach_{i,t}$)	-0.0808	0.0155 ***
Coach Lagged 1 Year ($Coach_{i,(t-1)}$)	-0.0066	0.0160
Draft Squared ($DraftSqr_{i,t}$)	0.00005	0.00004
Draft Squared Lagged 1 Year ($DraftSqr_{i,(t-1)}$)	0.00006	0.00004
Metro Size ($MetroSize_i$)	-0.0011	0.0012
State Income Tax Lagged 1 Year ($StateIncTax_{i,(t-1)}$)	-0.4167	0.2167 **
Observation	281	
R-squared	0.4612	
F(18,262), Prob>F	12.46	0.0000 ***

**Significant at the 5% level, ***Significant at the 1% level
* Significant at the 10% level

When comparing both regressions, the winning percentage model is what corresponds with that from Zimmer's paper. Much like his paper, my model has a positive coefficient for the lagged win percentage and negative coefficients for coaching change and top marginal tax bracket. I see an even higher strong negative coefficient for the tax bracket variable in my model, showing that higher tax brackets do have an inverse effect on win percentages and that teams in states with higher tax brackets are negatively affected. The coaching change variable is negative in both models, which means that a coaching change has a negative short-term effect on a team.

My constant of regression is statistically significant at a low alpha level, which is slightly problematic. It may have caused some of the discrepancies I see between the replication and the original model.

Interestingly, I didn't find win percentages lagged by two and three years to be statistically significant in my model. This may just be due to collecting and formatting errors within the replication.

Draft Picks had no effect on win percentage in the replication and in the original model, which is partly why I decided to add a draft pick trade binary variable in the second set of regressions.

Oddly, draft pick was statistically significant in the negative direction in my logit model, but it had almost no effect on the prediction with a coefficient of -0.00004. Tax bracket was also significant in the logit model, but at a lower alpha level.

Results with the New Data:

The new models I ran look at data from the 2010 to 2019 seasons predicting the win percentages and the "Z", or the rounded win percentages in a logit model. The results are shown below.

	Dependent variable:	
	Winning Percentage	Z Model
	OLS	*logistic*
	(1)	(2)
Winning Percentage (lag 1)	0.832^{***}	11.684^{***}
	(0.142)	(3.860)
Winning Percentage (lag 2)	0.060	4.375^{*}
	(0.095)	(2.524)
Winning Percentage (lag 3)	-0.092	-1.198
	(0.057)	(1.269)
Coaching Change	-0.059^{***}	-1.087^{***}
	(0.016)	(0.335)
Coaching Change (lag 1)	0.010	0.113
	(0.016)	(0.357)
Pre-Draft Trade	-0.006	-0.079
	(0.016)	(0.362)
Draft Pick Weight Squared	0.0001^{*}	0.001
	(0.0001)	(0.002)
Draft Pick Weight Squared (lag 1)	0.00003	0.002
	(0.00004)	(0.001)
Market Size	-0.001	0.021
	(0.002)	(0.035)
Annual Population Change (lag 1)	-0.209	-8.750
	(0.611)	(13.303)
Tax Bracket (lag 1)	-0.206	-8.090^{*}
	(0.199)	(4.630)
Constant	0.083	-7.094^{***}
	(0.101)	(2.694)
Observations	290	290
R^2	0.482	
Adjusted R^2	0.443	
Log Likelihood		-135.028
Akaike Inf. Crit.		312.057
Residual Std. Error	0.115 (df = 269)	
F Statistic	12.494^{***} (df = 20; 269)	
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Glaringly, the win percentage model actually shows an insignificant estimate for tax bracket data on winning percentage, but the logit model with rounded wins shows tax bracket has a strong negative significance. The clear reason for this disparity is unknown but it makes sense intuitively. The result may not be significant in the win percentage model due to the rise in salary. With players making more money overall, players may be more inclined to “bite the bullet” on higher taxes. With that being said, teams with high tax brackets and losing records should also have a harder time attracting free agents to that team, because the player is neither winning games, nor keeping more money. Furthermore, players have a greater incentive to join a losing team with a lower top marginal tax bracket than a losing team with a higher top marginal tax bracket.

Neither added variable of Pre-Draft Trade nor lagged Annual Population Change turned out to be statistically significant. The former essentially confirms that the draft has negligible impact on short-term winning percentages. The latter implies that the growth rates of the city itself don’t have an impact on its NBA team winning games.

Coaching Change is statistically significant and negative in both models, similarly to the replication, which again shows that it is statistically insignificant in the short term.

The first winning percentage lag term is statistically significant in the positive direction, like the replication. Winning percentage in the previous year seems to be a good predictor of winning percentage in the future.

Conclusion:

Considering Zimmer’s model, my replication, and my additions, we can conclude that a higher tax bracket still has a negative impact on winning. In future collective bargaining agreement negotiations, the effect of a higher tax bracket should be an economic consideration for the league. Additionally, the league could implement a clause where only post-tax dollars count towards the NBA salary cap and each team has a unique top marginal tax bracket-adjusted salary cap.

I believe the results change when considering the predictive power of winning percentage lagged by two and three seasons. Zimmer cites a cyclical nature of the league as a whole, where "league parity starts to take into effect after two seasons and team winning percentage starts to the norm." Based on my findings, I would disagree with that claim. Over this decade, we have seen more advanced analytics being used to make basketball-related decisions. GMs such as Sam Hinkie dread being mediocre and start tanking processes lasting for multiple years in order to contend for a title. While Zimmer's sentiment still may hold, I did not find any statistical backing towards this claim.

Draft picks and market sizes have marginal effect on winning percentages in the short term. Even when including extra variables in the model that had to do with both, they were both shown to be statistically insignificant. In future models, these variables are probably best being excluded.

In conclusion, tax brackets have an indirect negative effect on winning percentages, even today. This paper can be expanded on, given the low R-squared and adjusted R-squared values, but we see a clear negative disparity for teams in states with higher tax brackets.

Bibliography:

Adams, L. (2020, November 10). *NBA - CBA Minimum Annual Salary Scale*. RealGM. https://basketball.realgm.com/nba/info/minimum_scale/.

Contributor. (n.d.). *Land of Basketball - NBA Regular Season Standings from 1997 to 2019*. LandOfBasketball.com, Information About NBA Players, Teams and Championships. https://www.landofbasketball.com/.

The Long Term Perspective on Metro Areas. Macrotrends. (n.d.). https://www.macrotrends.net/.

Loughead, K. (2013, April 1). *State Individual Income Tax Rates, 2000-2014*. Tax Foundation. https://taxfoundation.org/state-individual-income-tax-rates/.

Loughead, K. (2021, February 17). *State Individual Income Tax Rates and Brackets 2015-2021*. Tax Foundation. https://taxfoundation.org/publications/state-individual-income-tax-rates-and-brackets/#Prior.

Many Contributors. (2017, June 6). *NBA Regular Season 2000-2001 to 2018-2019*. Wikipedia. https://en.wikipedia.org/wiki/NBA.

Marousek, F. (Ed.). (2005). *Professional Basketball Transactions Draft Pick Archive 2000-2018*. Professional Basketball Transactions Archive. https://www.prosportstransactions.com/basketball.

Zimmer, T. (2011, April 19). *Implications of State Income Tax Policy on NBA Franchise Success* . Research Gate. https://www.researchgate.net/publication/304625054_Implications_of_State_Income_Tax_Policy_on_NBA_Franchise.